LACUS FORUM XXX

Language, Thought and Reality

University of Victoria
over the past couple of decades, a great deal of research has investigated the manner in which words are represented in the mind and how they are accessed. The primary appeal of this branch of psycholinguistics is that it seeks not only to provide insight into lexical representation and processing, but also insights into the fundamental characteristics of human mental architecture. Yet, the research enterprise seems to be characterized by a troublesome paradox: Although it seeks to uncover important generalizations concerning the nature of language and mind, the methods that it employs typically involve single-word processing in the visual modality under highly artificial conditions that do not appear generalizable to the conditions of normal language processing.

Our goal in this paper is to explore this paradox by focusing on the question of whether the dominant experimental technique in the field, lexical decision, can indeed offer true generalizations concerning lexical representation and processing. We argue that this is fundamentally a question of ecological validity—the extent to which an investigation captures a phenomenon as it naturally occurs. Strictly speaking, the lexical decision paradigm does not meet this criterion. It typically involves the presentation of words and non-words in isolation on a computer screen under conditions in which the participant is not engaged in communicative activity, but rather is required to judge whether stimuli presented on the computer screen are, in fact, real words of the language. Although lexical decision is clearly not an example of normal language use, we present evidence that it might nevertheless meet the criteria of ecological validity because, although it tests language under artificial conditions, it also generates knowledge that has consequences for language processing under multiple situations.

The evidence that we present in support of this view comes from a comparison of data obtained using a classic lexical decision paradigm with data obtained from a new experimental paradigm that we have developed. Crucially, in this paradigm, target words are not presented in isolation, but are rather embedded in connected text under conditions in which the participant is instructed to read for story comprehension. Our initial results suggest a strong correspondence between the lexical processing of words in isolation and words in a story context. They also suggest that the new research paradigm may open up new lines of investigation, which we discuss at the conclusion of this paper.
1. **Validity as the Evaluation Metric of Science.** All investigators who employ an experimental approach to the investigation of language are aware that the adequacy of their research can be assessed under the general headings of reliability and validity. Of these two, the assessment of reliability is by far the more straightforward. Whether an experiment produces reliable results can be investigated directly through replication studies or, more commonly, indirectly by statistical means. The statistical analyses that are typically applied to psycholinguistic research have as their goal the determination of whether the results obtained in a particular experiment would also be obtained under identical conditions with other participants sampled from the same population (the analysis by subjects, or $F_1$) and with other language stimuli drawn from the same population of items (the analysis by items, or $F_2$).

The assessment of validity, the extent to which an experiment tests what it actually claims to test, is considerably more complicated and less objective. It is probably for this reason that the concept of validity has been traditionally decomposed into a number of discrete, but interacting components. We briefly overview four of the most commonly discussed aspects of validity.

The first is **content validity,** the extent to which an experiment adequately samples the population(s) under investigation. In the case of human participants, this translates into the extent to which the people who participate in the experiment constitute a representative sample of the population to which the research is designed to generalize. In the case of language structures, the criterion of content validity is often difficult to meet. Most experiments seek to learn something about “words in the mind”, yet, the practical constraints of experimentation require that only one or two languages are tested and that, within those languages, control procedures often disallow representative sampling of words in the language as a whole.

Whether an experimental investigation meets the criteria of the second type of validity, **construct validity,** can be the most contentious and not surprisingly, the least objective. Construct validity refers to the extent to which research tests theoretical constructs that can be shown to be relevant by virtue of empirical evidence or explanatory power. In the domain of linguistic inquiry, where researchers are often very divided on whether any set of putative principles or constructs actually exist, it is rarely the case that the criteria of construct validity can be met to everyone’s satisfaction.

The third type of validity, **face validity,** is defined as the extent to which a study has the appearance of a true experiment. Although face validity has often been dismissed as a false criterion of validity, it can be a very powerful force in shaping how research disciplines emerge and develop along methodological lines. In a recent review of methodological trends in mental lexicon research, Libben and Jarema (2002) surveyed 58 studies of lexical representation and processing, and found that 43% of these investigations employed the lexical decision task. In all but one of these investigations, response latency was the variable measured. The clear dominance of this research paradigm in the field has had the effect of imbuing lexical decision experiments with the aura of methodological prototypicality within the psycholinguistic research community. As such, it is perhaps optimal to view face validity as a type of ‘cultural valid-
Ecological validity, lexical decision, and lexical processing

ity’, i.e., validity that is both culturally defined and culture-specific. Within the culture of psycholinguistic investigation, lexical decision has become almost synonymous with lexical processing. But is this truly the case, or does it only appear to be so within the culture of a small research community? To address this question, we turn to the fourth and final type of validity which is at the core of our investigation.

1.1. Ecological validity and the lexical decision task. As noted above, the concept of ecological validity is tied to the concept of generalizability. An experiment is ecologically valid if it yields results that can be generalized to provide insight into phenomena as they occur in a natural (usually broader) environment. Thus, under this conceptualization, field research has intrinsic ecological validity, whereas laboratory psycholinguistic research most often needs to make the case for ecological validity. This, of course, is not unknown to experimentalists but is rather the result of a trade-off between the advantages of observation in a natural task setting and the advantages of control over tasks and stimuli.

In this trade-off, the lexical decision task certainly has substantial advantages in terms of control. It allows an experimenter to manipulate exactly what a participant will see, in what context, and for how long. For example, a researcher interested in whether the frequency of a particular word affects the ease (and therefore, speed) with which a word is recognized might select high and low frequency words for presentation, measuring the speed with which the ‘yes’ lexical decision response is made. Variations on this basic experimental design may involve manipulating the context of presentation so that stimulus words are preceded by related and unrelated stimuli (a primed vs. unprimed lexical decision task) to measure how words facilitate each other’s recognition. Crucially, the primes in such an experiment can be presented for any duration, including very brief periods (e.g. 40 milliseconds), which are sufficient for recognition but too brief to be consciously perceived.

Finally, it should also be noted that in lexical decision tasks, both real words and non-words may constitute the critical stimuli. Thus, an experiment that targets the effects of phonotactic (or orthotactic) constraints in visual processing may manipulate the orthographic properties of non-words to investigate whether strings such as ‘gloor’, which correspond to phonotactically legal strings in English, are rejected more slowly than strings such as ‘gmoor’, because the latter are less word-like.

We have alluded to the view that the dominance of the lexical decision task in psycholinguistic research is partially due to social factors that favour methodological cohesion within a research community. This, however, cannot be the main reason for its dominance. Lexical decision also offers researchers some real methodological advantages. The first of these is ease of use. Lexical decision tasks are relatively easy to create and require minimal laboratory hardware beyond a desktop or laptop computer. Analysis is relatively simple, because responses are discrete (‘yes’ or ‘no’). The measurement of response latency for each of these response types ensures a relatively sensitive dependent variable that is not subject to the floor or ceiling effects that often characterize accuracy measurements. Finally and most importantly, the lexical
decision paradigm is understood to provide a ‘pure’ measure of lexical recognition—one that simply measures how long it takes for a word to be initially accessed.

But the question remains: whatever its laboratory advantages, does this paradigm allow us to learn about lexical processing in general? In order to evaluate this question from the perspective of ecological validity, we might focus on two considerations. The first is the issue of experimental artifacts. If an experiment reveals a stable (i.e. reliable) pattern of behaviour that is, however, an artifact of isolated word processing, ecological validity is almost certainly compromised. On the other hand, if the results are not artifactual, we should not be led astray by the ‘face validity of ecological validity’. Put another way, it is not necessarily the case that just because a lexical decision task does not have the appearance of ecological validity, it does not yield results that are in fact generalizable to more natural contexts.

2. Once upon a lexical decision task. Our goal in the research reported here was to investigate the issues of ecological validity discussed above through the creation of a new experimental paradigm that would have some, but not all, of the characteristics of the classical lexical decision task. More specifically, we asked the question: are lexical decision results artifacts of the manner in which lexical decision tasks are normally conducted, i.e. the presentation of words in isolation rather than in connected text. It is quite conceivable, for example, that effects such as lexical frequency, as discussed above, are only obtained because, when words are presented in isolation, only lexical variables get to play a role. What would happen, for example, if participants were attending to a story instead? Under such conditions, it is conceivable that the frequency effect would simply disappear, because participants could use top-down processing to predict which words would be presented next. The effect could also disappear because in such a ‘natural’ context, the processing emphasis is on properties of the story, not on properties of words within it.

We sought, therefore, to construct a lexical decision task in a story context and to arrange that experimental context so that participants were required to pay attention to properties of the story by, for example, answering comprehension questions throughout the experimental session. This goal, however, led to our greatest design challenge: by definition, lexical decision experiments require the presence of real words and non-words for choice tasks. What type of story could contain the required large number of non-words, without itself sacrificing ecological validity as a natural story? The selection of fairy tales as a literary genre seemed to offer us a solution to this problem. Fairy tales often involve fantastic settings with novel names for characters, objects, and places. Our approach to the paradigm design capitalized on this property by embedding both words and non-words at natural points within fairy tale constructed for this experimental purpose. This fairy tale is presented in the appendix, with target words (i.e., those used for lexical decision) shown in bold.

As explained below, the 750-word fairy tale contained 34 target real words and 35 target non-words. Approximately half of the real words were high frequency and half were low frequency. Of the 35 non-words, approximately half were orthographically
and phonologically legal and half contained pairs of consonants that violated the phonotactic and orthotactic constraints of English. Thus, taken together, the critical words in the fairy tale story allowed us to test for both a real-word frequency effect and a non-word phonotactic legality effect in lexical processing.

Our investigation of the frequency and legality effects proceeded in the following manner: we extracted the critical words in the fairy tale and presented them to participants in a classic lexical decision task, in which words are presented one at a time in the center of a computer screen. This investigation is reported in Section 3 below. Following this experiment, a second group of participants were presented with the same list of words, now embedded in the fairy tale. The story was presented in the center of the screen one word at a time, for a duration of one second per word. Participants were required to attend to the content of the story, but were also asked to judge the lexicality of target words (presented in red) as they appeared in the story. The results of this second experiment and their comparability to those of Experiment 1 are presented in Section 4 of this report.

2. EXPERIMENT 1: CLASSICAL LEXICAL DECISION. One of the most robust effects in lexical decision experiments is the frequency effect. The frequency of a word has been found to be perhaps the strongest determinant of the speed with which a word is recognized, with high frequency words having an advantage over low frequency words. The source of this effect has been characterized in a variety of architecturally distinct models. Forster (1976) captured the frequency effect within the context of a lexical search model in which words in the mental lexicon can be conceived as being represented in a frequency-ordered list. A strongly contrasting view was represented in Morton’s logogen model (Morton 1969), in which high frequency words were seen to have low activation thresholds, so that they could be more easily activated than low frequency words under conditions of equal stimulation from the outside world. Currently, the logogen-type view of frequency can be said to dominate (with substantial refinements).

Another well-known effect in the lexical processing literature is that words that violate the phonotactic and orthotactic constraints of English (e.g. ‘gmoor’) are more easily rejected in lexical decision as compared with legal strings (e.g., ‘gloor’) that could conceivably represent real words (Libben 2000). The reason for this is likely related to depth of processing. Illegal non-words can be rejected out of hand, because they could not possibly exist in the language. In a Forster-like search model, legal words initiate an exhaustive search of the mental lexicon, resulting in long response latencies, because no corresponding entry is found in the participant’s mental lexicon. Activation models can predict the same results, but through different means. In activation models, illegal non-words excite no similar representations in the mental lexicon, because, by definition, none exist. Legal words, on the other hand, have at least some orthographic neighbours, which are automatically activated and then deactivated, thus increasing the time required to make a ‘no’ lexical decision.
In the experiment detailed below, our goal was to replicate each of these effects in a classical lexical decision paradigm, so that the obtained results could be compared to those found using the fairy tale paradigm in the same laboratory, using the same experimental software, and with participants drawn from a single participant pool.

3.1. METHOD.

3.1.1 PARTICIPANTS. Twenty undergraduate students from the University of Alberta participated in this experiment. Participants were between the ages of 18 and 30, and all were native speakers of English. Each was paid ten dollars for his/her participation.

3.1.2 PROCEDURE. Participants were tested one at a time in psycholinguistic testing booths. The experiment was conducted on iMac g3 computers using Psycscope 1.2 experimental software. The experimental session was conducted in under ten minutes and consisted of three blocks. The first was an instruction block in which participants received standard lexical decision instructions, asking them to press the ‘yes’ key if the word presented on the screen was a real English word. If the presented string was not an English word, they were instructed to press the ‘no’ key. Participants were told that both accuracy and latency were being measured, so they should try to respond as quickly and as accurately as possible.

Following the instruction block, participants completed ten practice trials, and were then asked whether they were ready to proceed to the main part of the experiment. This main experimental block consisted of 80 trials in which 40 real words and 40 non-words were presented in random order. Each trial began with the presentation of a fixation point on the screen for 500 milliseconds, followed by the presentation of the stimulus string. The stimulus remained on the screen until the participant pressed either the ‘yes’ or ‘no’ key. The pressing of the response key initiated the onset of the next trial.

3.2. RESULTS AND CONCLUSIONS. Of the 40 real words and 40 non-words in the experiment, only those 69 critical stimuli from the fairy tale paradigm were analyzed. All responses that were greater than 1200 milliseconds (4% of the data) were removed from the data set.

Response latencies to words and non-words were analyzed in separate analyses of variance by participants, with word type as the repeated measure.

As can be seen in Figure 1, the expected frequency effects were obtained, such that participants responded ‘yes’ to high frequency words (mean frequency = 315 per million) significantly more quickly (F(1,19) = 42.5, p < .0001) than they responded to low frequency words (mean frequency = 3 per million). The results for legal and illegal non-words also corresponded to expectations. As can be seen in Figure 1, participants rejected legal non-words (rt = 673 ms) more slowly than they rejected illegal non-words (rt = 595 ms). This difference was statistically significant (F(1,19) = 78.2, p < .0001).

In summary, this initial classical lexical decision experiment found both significant frequency effects for real word stimuli as well as significant phonotactic legality
effects for non-word stimuli. The next step in our investigation was to determine whether comparable effects would obtain under conditions in which words and non-words were presented in the context of a story. This second experiment is reported in Section 4 below.

4. EXPERIMENT 2: LEXICAL DECISION IN A FAIRY TALE CONTEXT.

4.1. METHOD.

4.1.1. PARTICIPANTS. The twenty participants in this experiment were drawn from the same participant pool as those who took part in Experiment 1. All were undergraduate students from the University of Alberta between the ages of 18 and 30, and all were native speakers of English. Each was paid ten dollars for his/her participation.

4.1.2. PROCEDURE. The stimuli that participants responded to in this experiment were identical to those used in Experiment 1. The difference between the two experiments lay in the context of presentation. In this experiment, participants were instructed that they would be presented with a story, one word at a time. They were asked to attend to the content of the story, as there would be three sets of comprehension questions presented at certain points during the experiment. They were also told that if a word of the story appeared in red print, they were to judge, as quickly and as accurately as possible, whether that word was an English word by pressing either the ‘yes’ or the ‘no’ response key.

The experiment required approximately thirty minutes to complete. Following the instruction block, participants were presented with a short practice story, which was followed by the main 750-word fairy tale. This fairy tale was presented in three blocks.

![Figure 1. Response latencies to words and non-words in classical lexical decision.](image-url)
of approximately 250 words with five multiple choice comprehension questions at the end of each block. Within each story block, words appeared automatically on the screen at one second intervals.

4.2. Results. As in Experiment 1, response latencies greater than 1200 milliseconds were removed from the data set, and latencies to words and non-words were analyzed in separate analyses of variance by subjects, with stimulus type as the repeated measure. As can be seen in Figure 2, the pattern of results obtained in this experiment was almost identical to that obtained in Experiment 1. There was a significant frequency effect for real words (F(1,19) = 46.5, p < .001), as well as a significant legality effect for non-words (F(1,19) = 66.7, p < .001). Although the two experiments showed an extraordinarily similar data pattern, it should be noted that lexical decision latencies were, on average, about 100 milliseconds slower in the fairy tale paradigm. We interpret this result to represent the fact that in this paradigm, not every word required a response. Thus it is likely, that lexical decision latencies across all stimulus types are built upon a constant ‘response shift’ latency that requires approximately 100 milliseconds.

5. General Discussion. We began this investigation by highlighting two key characteristics of research on lexical processing. The first is that investigations seek to gain insight into the fundamental characteristics of lexical processing and the organization of words in the mind. The second is that research in this field shows a dominant use of the lexical decision paradigm which, on the surface, is highly artificial. The question we sought to address was whether the effects obtained in lexical
decision paradigms artifacts of the presentation of words outside any textual context? In other words, do lexical decision tasks have the ecological validity that would be required for valid generalizations concerning human lexical processing?

We reported two experiments. The first employed a classical lexical decision task with high and low frequency words as well as legal and illegal non-words. The second experiment employed a new paradigm that we have developed. In this paradigm, the lexical decision task is embedded in a textual context, specifically a fairy tale. This text genre was selected because it licenses the presence of both words and non-words as part of the text.

Results from these two experiments were virtually identical, providing evidence that frequency and phonotactic legality effects are not artifacts of isolate word presentation out of context. In our view, this pattern of results across experiments has two important implications.

The first of these implications concerns the nature of lexical processing. We interpret the consistency of effects across text-independent and text-embedded contexts to reflect a computational encapsulation of lexical processing. Under this view, the properties of lexical processing that generate the frequency and legality effects that we found reflect automatic and obligatory processes of lexical access that are stable across contexts because they are encapsulated as subsystems within the overall cognitive system.

The second implication concerns the experimental opportunities that are created, if indeed lexical processing is identical in text-embedded and text-independent lexical decision. In our view, the fairy tale paradigm opens up opportunities to investigate lexical processing phenomena that have thus far been outside the scope of lexical decision research. For example, it has not been possible in the past to investigate whether frequency effects can be modulated by participants’ perceptions of who is actually producing the words to be judged. In principle, it is possible that frequency thresholds would be altered if participants perceived the story producer to be a child rather than an adult, or a non-native speaker rather than a native speaker of the language. We are currently investigating these possibilities by extending the fairy tale paradigm to the next step of naturalness—one in which videos of speakers accompany the presentation of the story and the embedded lexical decision task. In this way, the new paradigm will allow us to investigate lexical processing not only in a text-embedded context, but also in a socially embedded one.

REFERENCES


APPENDIX:
FAIRY TALE: THE PRINCESS AND THE PANDA

Once upon a time there lived a young princess named Wan. She loved animals. Her castle looked more like a zoo than a forest. Everyday Wan would walk in the forest with her giant panda named Ralph. They would look at the rwoos, swim in the lagoon and feed the birds. One day, they stopped to eat a bowl of rice and pekelom at their favorite tbolod. They did not hear the evil sorcerer sneak up from behind. Now this sorcerer was part man and part prass. He had the head of a homosapien but his arms, legs and prools were very large and covered with hair and kmouls. He grabbed the panda with one of his talons and started to carry him away. He yelled back that he was going to hold Ralph hostage, and probably eat him, unless Wan brought him the eye of Hraw by dusk. Poor Wan! There was no way that she could climb Mount Doom and get the eye of Hraw in one day. She would have to go ask prince Twan for help. Wan did not like Twan at all. She thought that he was a plogant and a snob. But unfortunately, he was the only one who could help. An hour later, Wan was sitting in Twan’s banquet hall explaining what had happened. When she was done, Twan groosed at her and said he never liked the stinky panda anyway. Wan was very mad, she grabbed a nearby npob and almost threw it at Twan’s head before he said he was only joking and would help her find old stinky. Wan and Twan packed their flom-dons and went to the dugout where Twan kept his lwosarg. It had grown since the last time Wan had seen it. Its wings were as big as julks and its beak had become as sharp as a tmoop. It was a mean looking thing. A few bjoplons later, they were flying over glaciers, coming close to Mount Doom. The eye of Hraw was protected by the scariest creature in the area, the giant ice platypus. Lucky for Wan and Twan, the platypus had become lazy and fat over the years because he only ate gopls. They landed the lwosarg on a bamboo breeb in front of the platypus’s cave and crept in the sgoib. The cave stank of rotten food and sloog. The platypus was asleep on his back, snoring loudly, with the eye of Hraw placed on his tummy. Twan attached a hook to the eye of Hraw and lifted it off the platypus. The platypus grunted and his fdeew began to shake. Wan and Twan spun around and ran as fast as they could, tripping over trouds and boutrs. They jumped onto the lwosarg and could hear the platypus’s vorps hitting the ground behind them. The lwosarg took off and they left the platypus stomping on the mountain ledge. The sun was starting to set and the lwosarg was flying quickly to the sorcerer’s house. As they were flying, Wan offered Twan a bite of her wbiot. She was actually starting to like him. The sorcerer’s house was next to a large plut. The lwosarg landed behind a tree and Wan and Twan walked slowly to the front door of the ylop. Wan could feel small tremors running through her hands as
she thought about the horrible things that the sorcerer might have done to her panda. They could hear noises and blups coming from inside the main qyit of the house. Wan lay down and peeked through the hort under the door. She could see the sorcerer and Ralph sitting on a sedb playing cards. What in the name of ferd is going on? she asked out loud. Twan was already banging on the side of the brog. The sorcerer opened the door Wan showed the eye of Hraw and demanded to have her panda back. said Wan. The sorcerer shook his whiskers and said that Ralph was his dost now. He did not want to be alone again. Wan’s mipn turned red but Twan yelled: that if you have to kidnap your friends and hold them as nopks, they’re not really your friends. Maybe if you asked Wan nicely she would let you come and visit Ralph at her quarters. Wan thought about it and then said she guessed it would be okay, as long as he didn’t act like a coutx. Wan and Twan stayed and had dinner with the sorcerer.