Journal of Neurolinguistics 51 (2019) 297–308

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Semantic processing of metaphor: A case-study of deep dyslexia

Hamad Al-Azary[a,*], Tara McAuley[b], Lori Buchanan[b], Albert N. Katz[c]

[a] University of Alberta, Canada
[b] University of Windsor, Canada
[c] University of Western Ontario, Canada

ARTICLE INFO

Keywords:
Deep Dyslexia
Semantic Processing
Metaphor

ABSTRACT

Deep dyslexia is characterized by the production of semantic errors (e.g., reading the word *weird* aloud as *odd*) during oral reading and greater difficulty reading aloud abstract words than concrete words. In this paper, we examine whether deep dyslexia affects higher-order semantic processing; namely, metaphor comprehension. To that end, we asked GL, a participant with deep dylexia, to rate novel metaphors (e.g., *language is a bridge*) for comprehensibility. The topics of the metaphors (e.g., "*language*" in the item above) varied on concreteness, such that they were either abstract or concrete. Also, the semantic neighbourhood density (SND) of the constituent nouns (e.g., "*language*", "*bridge*") was manipulated. In addition to metaphors, GL also rated literal (e.g., *a gorilla is an ape*) and semantically anomalous sentences (e.g., *arrival is a shoestring*). GL rated the literal sentences as maximally comprehensible and he also rated the abstract-low SND metaphors as comprehensible. However, other, more semantically rich metaphors (i.e., those with concrete constituents or high-SND constituents) were treated as non-comprehensible with ratings indistinguishable from ratings given to anomalous sentences. We discuss how GL's data align with models of deep dyslexia and metaphor processing.

1. Introduction

Deep dyslexia is an acquired language disorder characterized by a number of difficulties associated with oral reading (Coltheart, Patterson, & Marshall, 1987) with the most diagnostic of these difficulties being the production of semantic errors, such as reading the word *weird* aloud as *odd*. Another characteristic of the disorder is the presence of a concreteness effect, in which concrete words, such as *knife*, are read aloud with more success than abstract words, such as *guilt* (Coltheart et al., 1987). Although studying the semantic irregularities inherent in this disorder is valuable for understanding basic reading processes, it can, in addition, be of potential value for understanding higher-level processes such as metaphor comprehension. In this paper we first review studies of deep dyslexia and their implications for metaphor comprehension in this population. Second, we review issues with testing metaphor comprehension in people with deep dyslexia, who are typically aphasic. Third, we present the performance of a participant with deep dyslexia on a metaphor comprehension task. Finally, we interpret the data in the context of established models of deep dyslexia and of metaphor comprehension.

1.1. Deep dyslexia and semantic processing

Despite producing semantic errors during oral reading tasks, evidence suggests that people with deep dyslexia have access to intact semantic representations. For instance, they perform well in auditory word-association tasks (i.e., producing semantic

* Corresponding author.
E-mail address: alazary@ualberta.ca (H. Al-Azary).

https://doi.org/10.1016/j.jneuroling.2019.04.003
Received 8 November 2018; Received in revised form 8 March 2019; Accepted 23 April 2019
Available online 10 May 2019
0911-6044/

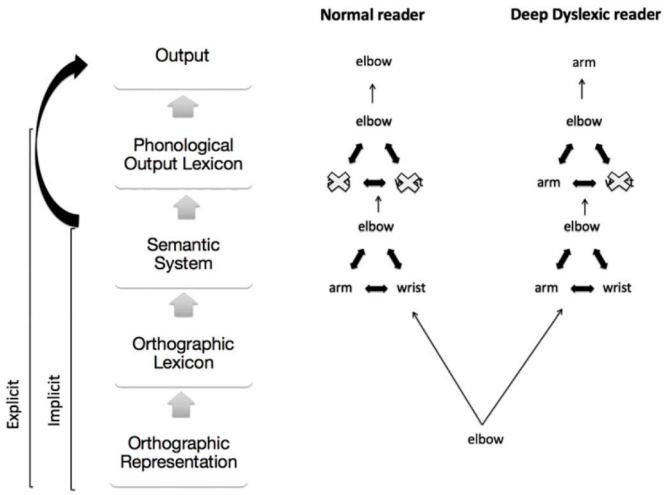

Fig. 1. Schematic diagram of a normal reader and a deep dyslexic reader, according to FIT. This diagram is sourced from Malhi, McAuley, Lansue, & Buchanan, 2018.

associates to spoken words) suggesting that semantic representations can be accessed when the task does not involve reading aloud (Colangelo, Stephenson, Westbury, & Buchanan, 2003). Furthermore, some reading studies demonstrate that spreading activation occurs within an intact semantic lexicon. For instance, JO, a participant with deep dyslexia, produced a greater number of semantic errors when words were presented in a block with other category exemplars (e.g., *jam, milk, flour, jelly,* etc.) than when words were blocked randomly with unrelated words, suggesting that spreading activation from the category exemplars interfered with reading aloud (Colangelo, Buchanan, & Westbury, 2004). In addition, people with deep dyslexia produce more semantic errors when the target word has many near semantic neighbours (i.e., is from a dense neighbourhood) than few near neighbours (i.e., is from a sparse neighbourhood), suggesting that activation is spreading to semantic neighbours within an intact semantic lexicon (Buchanan, Burgess, & Lund, 1996). Taken together, these studies indicate that people with deep dyslexia have a semantic system that is more intact and accessible than is suggested by their semantic reading errors (but see Riley & Thompson, 2010 for a counter argument).

As a critical component of their *failure-of-inhibition theory* (FIT), Buchanan, McEwan, Westbury, and Libben (2003) proposed that there is an intact semantic system in people with deep dyslexia. According to FIT, semantic errors are due to compromised inhibitory processes in the phonological output lexicon. In this view, when one reads a word, activation spreads to semantic neighbours of the target word and then to the corresponding representations in the phonological lexicon. For example, the written word *weird* will activate semantic neighbours such as *odd*. In non-impaired readers these semantic neighbours are inhibited in the phonological output lexicon (because that system also has input from orthography) and the target word can be read aloud successfully. Conversely, in a reader with deep dyslexia, semantic neighbours feed forward and remain activated in the phonological output lexicon; this failure of inhibition results in an increased likelihood of making a semantic error when reading aloud (see Fig. 1 for a schematic of FIT). Thus, in FIT it is proposed that the locus of impairment in deep dyslexia is at the level of the phonological output lexicon, whereas the semantic system remains intact. That is, for tasks that do not place any explicit demand on the phonological output lexicon (i.e., do not require reading aloud), we would expect someone with deep dyslexia to perform like a neurotypical reader. As such, FIT provides a basis for exploring metaphor processing in deep dyslexia, and this case-study is the first test of that.

1.2. Semantic processing of metaphor

Comprehending metaphors, such as "*that lawyer is a shark"*, involves activating the semantic representations of both the topic (i.e., *lawyer*) and vehicle (i.e., *shark*), and inhibiting semantic information unrelated to the metaphor's meaning (Black, 1955; Gernsbacher, Keysar, Robertson, & Werner, 2001). This process of activating and inhibiting semantic properties is detailed in Kintsch's (2000; 2008) predication algorithm. The predication algorithm is a computer model in which the meaning of sentences, such as *the horse runs,* is determined by searching the semantic neighbourhood of the predicate (i.e., *runs*) for properties related to the argument (i.e., *horse*) (Kintsch, 2001). The predication process is similar for modeling metaphoric meaning, where the topic and vehicle correspond to the argument and predicate of a literal sentence. However, for metaphors, the search for relevant properties involves more neighbours because the topic and vehicle tend to be semantically unrelated. Accordingly, to understand a metaphor such as "*that lawyer is a shark"*, the semantic neighbourhood of the vehicle must be searched for words that are also related to the topic (e.g., *vicious*); such words are factored into the metaphor's meaning. Moreover, words in the vehicle's semantic neighbourhood which are unrelated to the topic, and in turn, unrelated to the metaphor, are inhibited (e.g., *fish*). Therefore, metaphor comprehension results from both the activation and inhibition of semantic neighbours. Importantly, the predication algorithm simulates metaphor comprehension in a way that is consistent with human interpretation data (Kintsch & Bowles, 2002).

Both the topic and the vehicle of a metaphor can vary in their semantic content. For example, the topic and vehicle can differ in the number of close semantic neighbours, the so-called semantic neighbourhood density (SND). High-SND words have many near semantic neighbours whereas low-SND words have few (Buchanan, Westbury, & Burgess, 2001). Moreover, topics can vary in concreteness (whereas vehicles tend to be concrete). For instance, in the metaphor *language is a bridge*, the topic is abstract and the vehicle is concrete whereas in the metaphor *a pen is a sword*, both terms are concrete. Al-Azary and Buchanan (2017) studied how concreteness and SND may interact in metaphor comprehension tasks. They used metaphors that had either abstract or concrete topics whereas topics and vehicles were both either high or low-SND. This resulted in four semantic conditions, which, in descending order of semantic richness from high to low are (1) concrete-high SND (e.g., *a pen is a sword*); (2) abstract-high SND (e.g., *language is a bridge*); (3) concrete-low SND (e.g., *a pond is a mirror*); and (4) abstract-low SND (e.g., *responsibility is a chain*).[1] These metaphors were then used in comprehension tasks to assess what effects semantic density may have on comprehension and processing.

In an offline comprehension task, participants rated how much sense the items made (Al-Azary & Buchanan, 2017, Experiment 1). The low-SND (both abstract and concrete) metaphors were rated as the highest in comprehensibility, followed by the abstract-high SND metaphors. The novel metaphors constructed with arguably the semantically-richest items, concrete-high SND metaphors, were rated as the least comprehensible. To explain these results, Al-Azary and Buchanan (2017) proposed that semantic density is detrimental for metaphor comprehension: the many near semantic neighbours of the vehicle interfere with computing the new meaning for that item necessary in establishing the metaphor's meaning. The effects of topic concreteness were hypothesized as another source of semantic richness wherein the imaginable features of concrete topics must also be inhibited, to the extent that they are unrelated to a metaphor's meaning (e.g., the fact pens contain ink is irrelevant for *a pen is a sword*). In the case of low-SND metaphors, topic concreteness does not seem to play a role because there are fewer neighbours of the vehicle to inhibit and therefore, less of a burden on processing; in such cases, topic concreteness can be tolerated. Follow-up online comprehension tasks, (Al-Azary & Buchanan, Experiments 2 and 3) with the same items demonstrated that semantically dense (i.e., concrete-high SND) and sparse (low-SND and abstract-high SND) metaphors differ in their processing time courses. That is, semantically dense metaphors are equally comprehensible when presented for short (600 ms) or long (1600 ms) deadlines, suggesting shallow processing whereas semantically sparse metaphors become more comprehensible at long deadlines, suggesting deeper, elaborated processing. The conclusion arising from these data was that the predication process may be disrupted by rich semantic representations.

2. Testing metaphor comprehension in deep dyslexia

Predictions arising from FIT, along with the metaphor processing literature (e.g., Al-Azary & Buchanan, 2017; Kintsch, 2008) provide a compelling justification for the examination of metaphor comprehension in participants with deep dyslexia. Despite this, there are virtually no studies on the topic (but see Nenonen, Niemi, & Laine, 2002 for a case-study on idiom processing in a participant with deep dyslexia). One reason for this is that there are few tasks that are suitable for measuring metaphor comprehension in people with deep dyslexia. A standard reading aloud test is not optimal for several reasons: reading aloud a metaphor is particularly difficult for participants with deep dyslexia because of the presence of abstract words and function words (e.g., *the*, *is*), which are notably difficult for this population to read aloud. Additionally, asking for overt interpretations of metaphors is also problematic given the speech production deficits (i.e., aphasia) typically found in this population. Below, we will review how novel testing methods used on non-neurotypical participants can circumvent these problems.

2.1. Nominal metaphor comprehension in non-neurotypical populations

Despite decades of research on metaphor comprehension in neuropsychological patients, there is little consensus on how to study these populations, leading to differences in findings (see Coulson, 2008 for a review). One reason for the discrepancies is that neuropsychological studies of metaphor comprehension vary widely in task, stimuli, and patient selection and therefore are susceptible to different confounds (Schmidt, Kranjec, Cardillo, & Chaterjee, 2010). As such, we will limit our review to tasks involving reading nominal metaphorical sentences for meaning, that is *A is B* metaphors, where both *A* and *B* refer to nouns. This form is the most studied and understood of metaphoric variants studied in psycholinguistics.

There are three relevant studies in the literature, all of which have patients reading metaphors and choosing a “correct” interpretation from a set of foil options. In Mancopes and Schultz (2008) study, a participant with aphasia was administered a test based on the Metaphor Comprehension Task (from the Montreal Evaluation of Communication scale). The task consisted of reading fourteen metaphoric statements (9 nominal, e.g., *my cousin is a fridge*; 5 adjectival, e.g., *Ricardo is a sweet man*) and, for each, choosing one of three interpretations (i.e., the correct metaphoric interpretation from two foils). Their results showed that the correct (i.e., metaphoric) interpretations, were chosen 50% of the time by the participant. In a second, more recent study, Ianni, Cardillo, McQuire and Chaterjee (2014) also employed a multiple-choice metaphor reading task consisting of metaphors (e.g., *the coffee was a caffeine*

[1] In this ranking of semantic richness, we consider both topic-concreteness and SND as sources of richness. Thus, (1) concrete-high SND metaphors contain a concrete-topic and high-SND constituent nouns, and are therefore the semantically richest. (2) Abstract-high SND metaphors also contain high-SND constituents, but their abstract topics make them less semantically rich than concrete-high SND metaphors. Similarly, (4) abstract-low SND metaphors are the semantically sparsest of the metaphors because the topic is abstract and the constituent nouns are low-SND. (3) Concrete-low SND metaphors also contain low-SND constituent nouns, but the concreteness of the topic results in them being semantically richer than the abstract-low SND metaphors.

bullet) and four options (the correct, figurative option: *energy jolt*; the literal meaning: *military ammunition*; the opposite of the metaphoric meaning: *soothing lullaby*; and an unrelated meaning: *funny teacher*), along with literal sentences (e.g., *the police evidence was a bullet*). Ianni, Cardillo, McQuire, and Chatterjee (2014) tested three participants (two with left-hemisphere damage [LHD] and one with right hemisphere damage [RHD]) who previously showed no language impairment using traditional aphasia tests. They found three distinct response profiles reflecting metaphor impairment. The RHD participant demonstrated impaired performance with both literal and metaphoric sentences whereas one LHD participant also showed impairment in both sentence types, but greater impairment on metaphors; the remaining LHD participant showed impairment on metaphors but normal-like scores on literal sentences (see Cardillo, McQuire and Chaterjee, 2018 for a comprehensive follow-up study). Taken together, the results of these studies suggest that metaphor comprehension depends on contributions from both hemispheres, with damage to either potentially resulting in some type of metaphor comprehension difficulty.

2.2. Methods for testing metaphor comprehension in brain damaged participants

Although the multiple-choice method described above does not require participants (who may be aphasic) to generate interpretations, and it provides the advantage of producing scores that are amenable to statistical analysis, it is difficult to interpret their results for two reasons. First, the multiple-choice approach rests on the assumption that people completely fail to comprehend the metaphor if they choose a foil, rather than the ostensibly correct interpretation. However, this may be a premature conclusion because none of the multiple-choice options are overt indications of incomprehension (e.g., such as an option that indexes failure to comprehend e.g., *I don't know)*. For example, consider Ianni et al.'s (2014) test item, *the coffee was a caffeine bullet*. If a participant selects the literal-foil, *military ammunition*, this would be a metaphor comprehension failure according to their scoring criteria. However, one can imagine how *military ammunition* could plausibly suggest metaphor comprehension; for example, an interpretation could be, "*coffee is my ammunition to fight off sleepiness*". Typically, people list multiple features, not just one, when they consider features relevant for describing a metaphor's meaning (Roncero & de Almeida, 2015; Utsumi, 2005). Furthermore, Ortony (1975) even went so far as to argue that the semantic properties relevant for metaphor comprehension are unnameable. Therefore, a test that defines metaphor comprehension based on selecting the single best property may be too conservative.

Second, the multiple-choice approach used to date includes only metaphors (Mancopes & Schultz, 2008) or metaphors and literal sentences (Ianni et al., 2014). Neither approach allows for the comparison between metaphors and semantically anomalous sentences. Without anomalous sentences to serve as baselines one cannot conclude that the metaphors used in such tests are actually treated as metaphors. That is, if a reader chooses a foil rather than the correct target that does not necessitate that the metaphor is completely anomalous for that reader.

To address the limitations of the multiple-choice method, in the present case-study, we opted to use the testing procedure in Al-Azary and Buchanan's (2017) first experiment, briefly reviewed above in section 1.2. This involves the presentation of three item types; namely, literal sentences (e.g., *a gorilla is an ape*), metaphors (e.g., *justice is a net*), and anomalous sentences (e.g., *veneration is a pickle*). Unlike the multiple-choice method, in our test, metaphor comprehension is not inferred from selecting a particular interpretation; rather, comprehension is inferred by comparing comprehensibility ratings of the items. That is, if metaphors are rated to be more comprehensible than anomalous sentences, we infer that metaphors were comprehended as metaphors. Furthermore, because the metaphors vary in their semantic representations, we can characterize comprehension more specifically than in previous tests. Recall that the participants in Ianni et al.'s (2014) and Mancopes and Schultz (2008) comprehended at least 50% of the metaphors on which they were tested, indicating that the ability is partially intact. However, because in those tests the nominal metaphors are not semantically manipulated, it is difficult to pinpoint the nature of the deficit. Our approach allows us to answer this question by using metaphors that differ on semantic dimensions and assessing which semantic conditions may be particularly difficult to comprehend.

Comprehensibility ratings, similar to those employed in our procedure, have been used in a number of metaphor processing studies (e.g., Carriedo, et al., 2016; Jacobs & Kinder, 2018; as reviewed by; Jones & Estes, 2006). As an offline measure, rating a metaphor does not capture the online time-course of processing, but reflects downstream processes such as elaboration that may be taken into account when making a rating (Gibbs & Colston, 2012). As such, this measure is comparable to the multiple-choice task but without the potential pitfalls reviewed above.

2.3. Present study and research aims

In the present case-study, GL rated printed literal, metaphoric, and anomalous statements for comprehensibility. Because this is the first study of metaphor comprehension with a participant with deep dyslexia, and since theories of the disorder are mostly limited to semantic processing in oral reading of single words, we do not have any firm predictions. Rather, we are interested in the question of whether GL would demonstrate sensitivity to the semantic differences underlying literal, metaphoric, and anomalous sentences. Moreover, we are interested in characterizing GL's response profile to the metaphors; would he rate the metaphors similar to neurotypical participants (as described in Al-Azary & Buchanan, 2017)[2]? If so, this would suggest his metaphor processing abilities are unimpaired. However, if GL's response profile is unique, then his data may be interpreted in light of the disrupted cognitive processes underlying deep dyslexia.

[2] As undergraduate students, the neuro-typical participants are younger and slightly less educated than GL. As such, this sample is to serve as a general comparison. The primary analysis will be based on an item-analysis of GL's ratings.

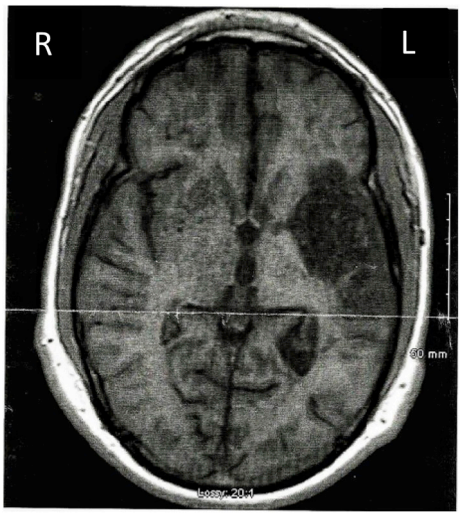

Fig. 2. CT Scan of GL following a large, left MCA stroke.

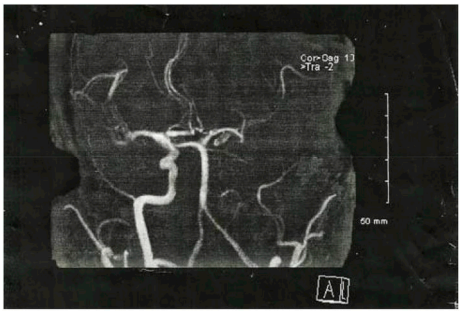

Fig. 3. Magnetic Resonance Angiogram (MGA) confirming the left internal carotid artery dissection.

3. Patient, demographics and stroke details

GL (at time of testing) is a 35-year old male with 17 years of formal education. At the age of 26, GL was diagnosed with a cerebrovascular accident (CVA) in the left middle cerebral artery (MCA) resulting in hemiparesis on the right side of his body and aphasia. Computed Tomography (CT) scans confirmed that the left MCA stroke (see Fig. 2) was secondary to a left internal carotid artery dissection (see Fig. 3). GL experienced a few complications following the CVA, including grand-mal seizures and post-stroke depression. GL has been reasonably healthy since then and has received speech therapy intermittently. He was paid $10 for participation in this study.

3.1. Speech-language assessment pre- and post-treatment

GL was assessed in a rehabilitation setting by a speech language pathologist via a standard aphasia protocol. In the initial assessment when entering rehab, GL met criteria for moderate to severe deficits in auditory and reading comprehension, as well as verbal and written expression. At that time, the hearing, swallowing, speech/voice, and cognitive (i.e., attention and concentration were only assessed) assessments were all determined to be within functional limits. GL was re-assessed prior to being discharged and improvements were noted in auditory and reading comprehension, and verbal and written expression. In particular, GL's responses to simple yes/no questions, ability to read letters and numbers, attention and concentration, immediate memory, and short-term memory were determined to be within normal functional limits. He still demonstrated a mild deficit in his ability to complete picture-word discrimination, follow one step commands, read single words, and repeat short strings of words. GL continued to demonstrate a moderate to severe deficit in understanding complex yes/no questions, following two or three step commands, understanding paragraphs, reading simple sentences and paragraphs, naming objects, producing sentences, and writing words. Post rehabilitation assessment indicated that GL's receptive and expressive language difficulties, perseveration on words and thoughts, and phonemic and verbal paraphasias persisted. However, since his release from the rehabilitation facility GL has continued to improve, and he has been a regular visitor to our laboratory for the past three years. His initial problems with receptive language (e.g., difficulty understanding complex instructions) appear to have resolved. Some of his expressive language difficulties persist, as he still has very profound word finding difficulties in spontaneous speech and his oral reading is limited to effortful production of single words.

Table 1
GL's reading error types.

Word Class	Target Word	Response	Type of Error
Adjective	Glorious	Glorify	Morphological
Noun	Slavery	Slavement	Morphological
Noun	Movement	Moving	Morphological
Noun	Laughter	Laughing	Morphological
Noun	Insanity	Insaneness	Morphological
Noun	Hatred	Hating	Morphological
Noun	Bravery	Bravement	Morphological
Verb	Marry	Marriage	Morphological
Verb	Activation	Action	Morphological
Adjective	Weird	Odd	Semantic
Adjective	Soft	Soap	Semantic[a]
Adjective	Major	Corporal	Semantic
Noun	Kindness	Likeness	Semantic
Noun	Grief	Hurt	Semantic
Noun	Tolerance	Honorable	Semantic
Verb	Remember	Understand	Semantic
Noun	Rage	Cage	Visual
Adjective	Sparse	Spark	Visual
Noun	Infancy	Infanty	Visual
Noun	Elegance	Element	Visual
Noun	Defeat	Define	Visual
Noun	Advantage	Arranged	Visual
Verb	Detect	Detach	Visual
Verb	Illustrate	Illusion	Visual
Verb	Exist	Exit	Visual
Adjective	Rare	Hair	Visual/Phonological

[a] This error is considered semantic because raters believed that it likely came from confusion with the Canadian Brand name Softsoap.

3.2. Results of word-reading tasks

GL was tested on word reading tasks to determine whether he produced responses similar to those reported in other case studies of people with deep dyslexia (e.g., Buchanan, McEwen, Westbury, & Libben, 2003; Colangelo et al., 2003). On multiple word-reading tasks consisting of different word classes, GL has made a variety of reading errors. See Table 1 for examples of reading errors made on these tasks. Based on this error profile, and in particular, the presence of the diagnostically necessary (and sufficient) semantic errors, GL meets criteria for deep dyslexia. In addition, GL made a number of morphological errors, which are also thought to characterize readers with deep dyslexia (Coltheart, 1981), and in English they are predominant on the right side of a target word. This might suggest that GL failed to follow instructions or have a visual field cut, but GL has performed numerous linguistic and non-linguistic (e.g., picture naming, matching tasks) experimental tasks in our lab, and he has never had difficulty with following complex instructions. Therefore, we are confident that GL's mistakes in reading are not because of a failure to comprehend task instructions. Furthermore, GL has never demonstrated any evidence of visual field neglect and he retains and uses a driver's license.

3.3. Materials

All items were sentences taken from Al-Azary and Buchanan's (2017) Experiment 1 and were of the form *A is B*, where *A* and *B* are both nouns. These consisted of literal (e.g., *a gorilla is an ape*), metaphorical (e.g., *a mosquito is a vampire*), and anomalous (e.g., *a fork is a planet*) sentences, which varied on semantic neighbourhood density (SND; derived from the WINDSORS model – see Durda & Buchanan, 2008) as well as topic concreteness. In brief, the (24) literal items are concrete statements, where the constituent nouns are

Table 2
Example of stimuli per semantic condition.

	Abstract-Concrete		Concrete-Concrete	
	High SND	Low SND	High SND	Low SND
Anomalous	Argument is a Paint	Shelter is a Nose	A Cake is a Wrench	A Circus is a Pool
Literal	–	–	Banana is a Fruit	Gasoline is a Fuel
Figurative	Passion is a Storm	Responsibility is a Chain	A Mosquito is a Vampire	A Cloud is a Curtain

both concrete. Of these literal items, 12 are high-SND, such that both constituent nouns have many near semantic neighbours and 12 are low-SND, such that both constituent nouns have few near semantic neighbours.

The metaphors (48) varied on topic-concreteness such that topics were either abstract or concrete (vehicles were always concrete). Moreover, SND was manipulated, such that both the topic and vehicle in a metaphor were either high or low-SND. For instance, a high-SND metaphor has a topic and vehicle both of which are high-SND words whereas a low-SND metaphor has a topic and vehicle in which both are low-SND words. This resulted in four semantic conditions, each containing 12 metaphors; concrete-high SND (e.g., *a pen is a sword*), abstract-high SND (e.g., *language is a bridge*), concrete-low SND (e.g., *a pond is a mirror*), and abstract-low SND (e.g., *responsibility is a chain*). Also included were 48 anomalous sentences (e.g., *veneration is a pickle*), which have the same constituent characteristics (i.e., concreteness and SND) as the metaphors and therefore fall under the same experimental conditions as the metaphors. However, these sentences were intentionally constructed to be meaningless. See Table 2 for an example of items (the entire item list can be found in Al-Azary & Buchanan, 2017).

3.4. Procedure

The procedure was the same used in Experiment 1 of Al-Azary and Buchanan (2017). The task was administered on a Dell PC with Windows XP operating system using Direct RT Software (Jarvis, 2012). Stimuli were presented in proper case in the center of the screen in size 24 Times New Roman bold-faced font.

The participant gave informed consent to participate in the task. The entire task was completed on a PC. The researcher read aloud the instructions to GL and he performed practice trials to ensure comprehension of the task. The researcher remained in the room throughout the task. GL used his left hand and the number pad to provide suitability ratings[3](from 1, being very low, to 6, being very high) that corresponded to the appropriate number. Each item was presented individually on the screen in a randomized order, and GL was encouraged not to read the sentences aloud. Halfway through the task, GL was allowed a break before continuing. The task took approximately 45 min to complete.

4. Results

GL's literal, metaphoric, and anomalous sentence data, alongside the participants' data from Experiment 1 of Al-Azary and Buchanan (2017) who performed the same task, are graphed in the following sections.

4.1. Literal sentences

Like the university dsample, GL performed at near ceiling for the literal statements. The low-SND literal statements were all rated at ceiling (i.e., 6). The high SND statements were near ceiling, but one item (i.e., *A beard is Hair*) was rated the lowest value, as a 1. Despite the low rating on this one item, GL's average ratings for the literal sentences were similar to the average ratings from the intact university sample (see Fig. 4). The near-ceiling effects obtained here suggest that GL was engaged in the task and a good faith interpretation of the results is that his ratings reflected comprehensibility.

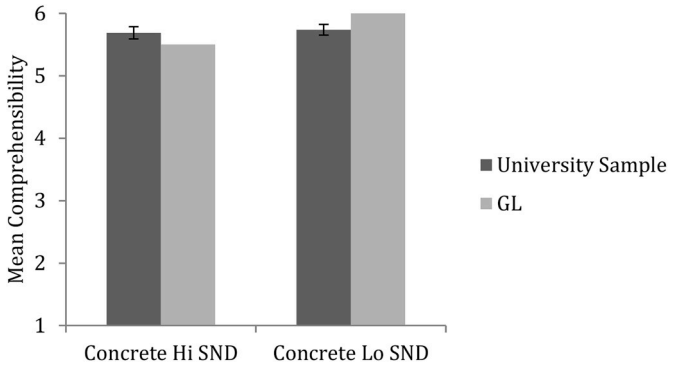

Fig. 4. GL's mean comprehensibility ratings for literal statements, alongside a neurotypical sample. Error bars represent the 95% confidence interval.

[3] For consistency with Experiment 1 of Al-Azary and Buchanan (2017), we asked GL to rate how suitable the sentences are. Thus 'suitability' is a proxy for comprehensibility. Such aptness ratings are highly correlated with comprehensibility ratings (Jones & Estes, 2006). Moreover, in other experiments using the same items, we found that comprehensibility ratings are virtually the same as the suitability ratings used here (see Al-Azary & Buchanan, 2017 [Experiments 2–3]; Katz & Al-Azary, 2017).

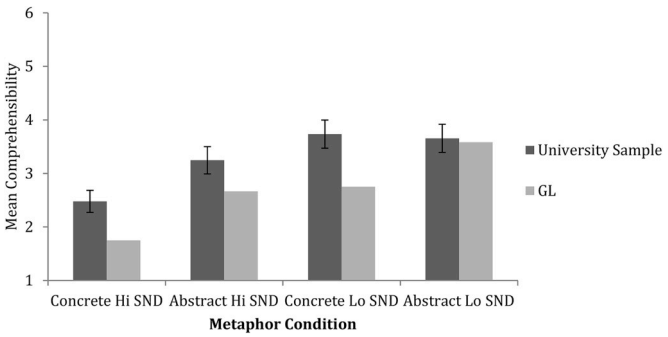

Fig. 5. GL's mean comprehensibility ratings for metaphors, alongside a normal control sample. Error bars represent the 95% confidence interval.

4.2. Metaphors

GL demonstrated sensitivity to the semantic richness of the metaphors in two ways. First, GL rated the concrete-high SND metaphors, which are the semantically richest, particularly low in comprehensibility (1.75) and within the range he rated anomalous sentences. Second, ratings given by GL for the abstract-low SND metaphors, which are the semantically sparsest, are rated the highest and are virtually identical to the data from the intact university sample (see Fig. 5). Because GL's data are difficult to interpret without reference to his anomalous sentence ratings, which were particularly high, we compared his ratings for the metaphors with his ratings for the anomalous items. Since the small number of items in each condition (12) results in low power for independent samples *t*-tests, we elected to perform a single sample *t*-test (one-tailed) for each metaphor condition. We used the mean of GL's anomalous statement ratings (2.08) as the hypothesized baseline value representing anomalous sentences.

Only the abstract-low SND items were rated as significantly more comprehensible than the anomalous sentences t (11), = 2.69, p = .01 by GL. His concrete-low SND metaphor ratings vs. the anomalous sentence ratings approached significance t (11), = 1.63, p = .07. The high-SND metaphors did not differ from anomalous sentences (abstract: p = .14, concrete: p = .82). Therefore, with the caveat that we are talking about a small sample of statements in any one condition, we conclude from the present data that GL found the abstract-low SND metaphors to be more comprehensible than anomalous statements and in general that the low SND items were more sensible than their high-SND counterparts for GL.

We also calculated how much higher GL rated the metaphors in a given semantic condition than the anomalies in the same condition (see Fig. 6). Concrete-high SND metaphors were rated only 10% more comprehensible than concrete-high SND anomalies. Similarly, abstract-high SND metaphors were rated 13.3% more comprehensible than abstract-high SND anomalies. Concrete-low SND metaphors were rated 27.6% more comprehensible than concrete-low SND anomalies. The biggest difference was observed with abstract-low SND metaphors, which were rated as 42.25% more comprehensible than abstract-low SND anomalies.

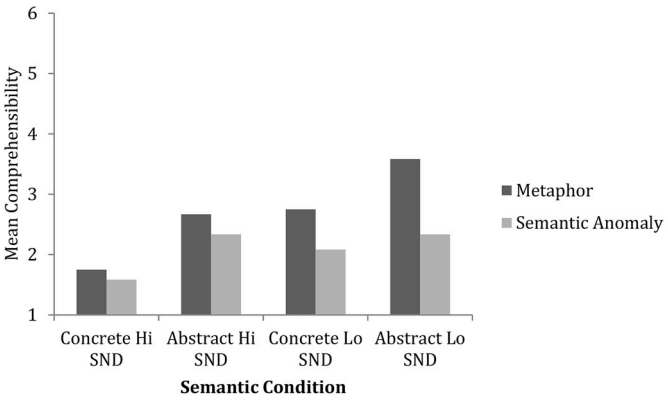

Fig. 6. GL's comprehensibility ratings for metaphors and semantic anomalies.

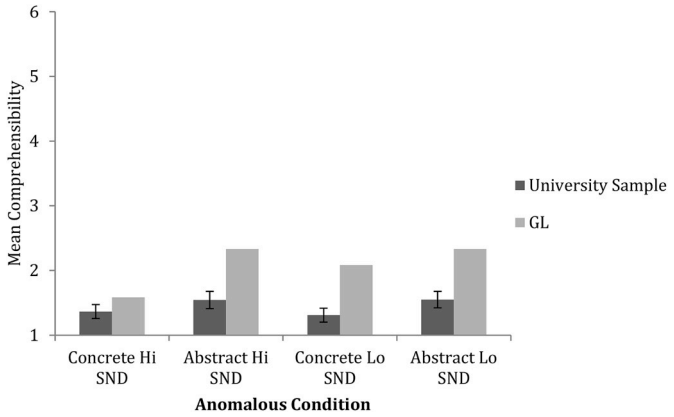

Fig. 7. GL's mean comprehensibility ratings for anomalies, alongside a normal control sample. Error bars represent the 95% confidence interval.

4.3. Anomalous sentences

Fig. 7 compares GL's ratings for the anomalous sentences with those of the university sample. Anomalies were rated, on average, lower than literals and metaphors. However, some anomalous items were rated very high – for example, *Art is a Kitten* and *Arrival is a Shoestring* were rated as maximally comprehensible (6), and items such as *a Trunk is a Gear, Depression is a Party,* and *a Toe is a Coach* received relatively high scores of 5. Clearly, some anomalous items were particularly meaningful for GL. However, the anomalous sentences, on average, were the lowest rated items which further indicates that GL's ratings reflect comprehensibility.

5. Discussion

Our objective was to determine whether the disrupted processes in deep dyslexia would affect metaphor comprehension. To that end, we asked GL to rate novel written metaphors, along with literal and anomalous sentences, for comprehensibility. GL rated the literal statements as being the most comprehensible and the anomalous sentences as being the least comprehensible, which indicates he was engaged in the task and his subjective ratings indeed reflected comprehensibility. Moreover, semantically sparse metaphors (i.e., abstract low-SND) were rated as comprehensible, relative to anomalous sentences, whereas semantically richer metaphors were rated as anomalous. To interpret our findings, we compare and contrast GL's response profile with two populations; namely, (1) with other brain-damaged participants who engaged in similar reading-based metaphor comprehension tasks, and (2) the intact university sample that rated the same items (Al-Azary & Buchanan, 2017 [Experiment 1]). Lastly, we will consider how GL's data aligns with established models of semantic processing.

With regards to other neuropsychological studies reporting metaphor comprehension difficulties with reading-based tasks, they do not describe why particular items are problematic for people with brain-lesions (Cardillo et al., 2018; Ianni et al., 2014; Mancopes & Schultz, 2008). Although such results are interesting, they do not provide a foothold in trying to determine why people with brain-lesions have comprehension difficulties. However, our data suggest a novel patient profile in which semantically rich metaphors in particular were problematic for GL. This provides a nuanced picture of GL's performance, demonstrating normal-like performance on both literal statements and metaphors of a particular semantic condition. Moreover, this fine-grain characterization allows us to isolate the comprehension deficits based on the semantic representations of the metaphors. Future research in this domain should therefore consider how semantic distinctions between metaphors' (e.g., concreteness, density) may affect their comprehensibility. Doing so may suggest that brain-damaged participants are not either impaired or unimpaired with regard to metaphor comprehension; rather, comprehension difficulties may vary on a continuum depending on item characteristics.

GL and the university students performed similarly on literal sentences, in which category membership statements were rated as maximally comprehensible. This finding in particular conforms to the prediction from FIT that people with deep dyslexia, such as GL, have a largely intact semantic system. However, GL's performance differed from the university sample in some ways. That is, he rated most of the metaphors as considerably less meaningful than did the university sample. Furthermore, GL's ratings for those metaphors did not significantly differ than his unusually high comprehensibility ratings for anomalous sentences. Whereas all the metaphor conditions are more comprehensible than anomalous sentences for the intact university sample, the same cannot be said for GL. For him only abstract-low SND metaphors are more comprehensible than anomalous sentences.

GL's selective comprehension of semantically sparse metaphors (abstract low-SND) provides some support for Al-Azary and Buchanan's (2017) semantic density hypothesis, which holds that semantic density is detrimental for metaphor comprehension in general. Recall that Al-Azary and Buchanan (2017) found that topic concreteness and SND interact in metaphor comprehension tasks, with low-SND abstract and concrete metaphors being equally comprehensible, followed by abstract high-SND and lastly, concrete high-SND metaphors (the control data in Fig. 5). They concluded that a metaphor's many near semantic neighbours disrupt processing of its meaning. In such cases when a metaphor has many near neighbours, topic concreteness becomes detrimental to processing as well. Although GL did not replicate this pattern in its entirety, he nonetheless showed sensitivity to semantic richness of

metaphor in a way that is consistent with the semantic density hypothesis. That is, he considered only the semantically sparse metaphors as comprehensible, which suggests such metaphors were advantaged compared to their semantically richer counterparts.

Beyond metaphors, GL demonstrated difficulties with the semantically rich items in general; the only literal sentence that GL rated as less than maximally comprehensible was a concrete-high SND sentence (which GL rated as minimally comprehensible, 1). Moreover, of the anomalous sentences, GL rated concrete-high SND sentences as the lowest. Therefore, the data we report supports the idea that semantic density is detrimental for processing metaphors, along with other argument-predicate sentences (such as the ones employed in our study) by showing that (1) semantically sparse metaphors (i.e., abstract-low SND items) are the most comprehensible for GL and (2) that semantically dense literal and anomalous items were generally more difficult to comprehend than similar items from other semantic conditions.

The Failure of Inhibition Theory provides a framework as to why high-SND sentences were generally rated as less comprehensible than low-SND sentences (of all three sentence types). Processing argument-predicate sentences (whether literal, metaphoric or anomalous) involves the activation of the constituents' semantic neighbours and the subsequent inhibition of any neighbours unrelated to the argument (Kintsch, 2001). Recall that in FIT, it is argued that semantic neighbours' representations in the phonological output lexicon are a particular source of interference for participants with deep dyslexia in reading-aloud tasks. In the current study, GL was not asked to read-aloud the sentences, yet his response profile demonstrated to us that he had difficulty comprehending items with many semantic neighbours. On the surface then, the findings might suggest that the evidence contradicts the FIT model. However, silently reading sentences requires that GL keep the topic and vehicle activated in his phonological loop (e.g., Baddeley, 2012) while making decisions about the comprehensibility of the statements. Reliance on the phonological loop, in turn, relies on an intact phonological working memory. Similar to a compromised phonological output lexicon, GL may have difficulty inhibiting semantic neighbours in his phonological working memory, a position compatible with FIT. Furthermore, failure to inhibit the phonological representations of the topic and vehicle's semantic neighbours may result in the re-activation of their semantic representations. As such, the reactivated semantic neighbours may disrupt metaphor processing. Thus, semantic effects may arise as a result of phonological working memory impairment and we will pursue this possibility more closely in follow-up studies.

Our results support this proposal as GL had more difficulty comprehending semantically rich sentences (i.e., high SND items) than semantically sparse sentences (i.e., low SND items). In particular, it can be argued that anomalous sentences recruit greater activation of semantic neighbours of the constituent words than other sentence types. That is, anomalous sentences, unlike metaphor and literal sentences, do not have many shared neighbours among their constituents. This results in searching a large semantic neighbourhood for potentially relevant neighbours (Kintsch & Bowles, 2002). As such, a relatively high number of neighbours may be activated, and due to an impairment in phonological working memory, remain activated and interfere with comprehension. The fact that high-SND anomalous sentences were rated as less comprehensible than low-SND counterparts supports this possibility because the former activates more irrelevant semantic neighbours than the latter. However, this is a post-hoc speculation which ought to be confirmed in future research.

In addition to a low-SND effect, GL demonstrated an abstractness effect such that abstract sentences (both metaphoric and anomalous) were rated as more comprehensible than concrete sentences. This aligns with another case study involving GL where abstract word pairs presented in an iconic order (*peace* presented spatially above *war*) were responded to faster than concrete word pairs (*desk* presented spatially above *carpet*) (Malhi et al., 2018). Moreover, it is also consistent with previous work on abstract word comprehension in deep dyslexia with participant LW (Newton & Barry, 1997). Despite only being able to read aloud 7.5% of abstract words, LW correctly recognized 90% of high frequency and 70% of low frequency abstract words in a lexical decision task. Moreover, LW demonstrated comprehension of many of the abstract words she was tested on, correctly matching pictures, synonyms, and definitions to abstract target words, and scored over 70% correct on these tasks (LW performed poorly on matching definitions to low-frequency abstract words). Therefore, LW demonstrated the ability of someone with deep dyslexia to understand some abstract concepts literally, whereas in the current case-study, GL demonstrated the ability to understand some abstract concepts metaphorically.

Malhi et al. (2018) also found that GL performed similarly to neurotypical participants on tasks that relied on implicit access to semantic representations, consistent with FIT (Buchanan et al., 2003). Contrary to their findings, and as reported above, here we found that GL diverged from intact readers in his comprehensibility ratings for anomalous sentences, along with the minimal rating for a literal sentence (*A beard is Hair*). A closer examination reveals key differences in the two studies methodology. First, Malhi et al. (2018) stimuli was comprised of word pairs, whereas sentence reading is inherently more demanding on the semantic and phonological systems. Second, and relatedly, Malhi et al. (2018) tasks were phonologically implicit, whereas a sentence reading task is phonologically explicit to the extent that the phonological working memory system must remain engaged during sentence comprehension (as described above).

Upon considering both FIT and the semantic density hypothesis, we explain the processes underlying GL's response profile as follows. In an average neurotypical person, semantic density, both from near neighbours and concreteness, disrupts metaphor processing. When metaphors are semantically dense, finding semantic properties which are appropriate for comprehension among many which are inappropriate becomes more difficult than for semantically sparse metaphors. Moreover, inhibiting those semantic properties which are unrelated to the metaphor's meaning is also particularly difficult in such a situation. This makes the metaphor less comprehensible than a semantically sparse metaphor, but still comprehensible enough to be considered as a metaphor rather than as a semantic anomaly. Conversely, for people with deep dyslexia inhibiting the high number of semantic neighbours (and their phonological representations) inherent in semantically dense metaphors is particularly challenging because of the reliance on an impaired phonological working memory system. As semantic neighbours remain activated in phonological working memory, their semantic representations may become reactivated as well. In such cases, the dense semantic representations are a persistent source of

interference during processing, resulting in heightened difficulty to find the semantic properties of the vehicle which are relevant to the metaphor and consequently leading the participant with deep dyslexia to treat semantically dense metaphors as anomalies. In contrast, semantically sparse metaphors (i.e., abstract-low SND) are comprehensible because they create less interference than their semantically dense counterparts. In this case, a reader with deep dyslexia processes such metaphors with little to no disruption, similar to an average neurotypical person.

We acknowledge that metaphor comprehension is a dynamic process, affected by both linguistic and extra-linguistic factors (Gibbs & Colston, 2012). Furthermore, metaphor comprehension can be assessed by numerous tasks, some of which are more ecologically valid than others (see, for instance, spoken metaphor comprehension tasks; Blasko & Connine, 1993; Chouinard, Volden, Hollinger & Cummine, 2018). Therefore, the impairment we report may be limited to our reading-based comprehension task. As such, future research ought to consider incorporating testing paradigms that reflect online communication processes in order to better characterize the nature of metaphor processing impairments reported here and elsewhere (e.g., Cardillio et al., 2018; Ianni et al., 2014; Mancopes & Schultz, 2008).

In summary, we explored here how a participant with deep dyslexia performed on a metaphor comprehension task. We have discussed how our results are consistent with established models of semantic processing. Moreover, because we studied a patient with a well-documented disorder and used items which varied on semantic dimensions, we were able to extend our understanding in this context from previous neuropsychological studies on metaphor comprehension. Future neuropsychology research on metaphor comprehension should therefore go beyond describing the locus of brain damage. Rather, researchers should consider the participants' disorder and how the resulting compromised processing mechanisms may disrupt comprehending particular stimuli. Doing so will better characterize metaphor processing and comprehension and inform theoretical accounts of semantic processing.

Acknowledgements

We are grateful for GL's willingness to participate in the ongoing research in LB's lab.

This research was supported by a Social Sciences and Humanities Research Council Words in the World Partnership Grant (895-2016-1008) awarded to LB

References

Al-Azary, H., & Buchanan, L. (2017). Novel metaphor comprehension: Semantic neighbourhood density interacts with concreteness. *Memory & Cognition, 45*(2), 296–307.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1–29.

Black, M. (1955). Metaphor. *Proceedings of the aristotelian society. Vol. 55. Proceedings of the aristotelian society* (pp. 273–294). Oxford, UK: Oxford University Press No. 1.

Blasko, D. G., & Connine, C. M. (1993). Effects of familiarity and aptness on metaphor processing. *Journal of Experimental Psychology Learning Memory and Cognition, 19*, 295–308.

Buchanan, L., Burgess, C., & Lund, K. (1996). Overcrowding in semantic neighborhoods: Modeling deep dyslexia. *Brain and Cognition, 32*(2), 111–114.

Buchanan, L., McEwen, S., Westbury, C., & Libben, G. (2003). Semantics and semantic errors: Implicit access to semantic information from words and nonwords in deep dyslexia. *Brain and Language, 84*(1), 65–83.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*(3), 531–544.

Cardillo, E. R., McQuire, M., & Chatterjee, A. (2018). Selective metaphor impairments after left, not right, hemisphere injury. *Frontiers in Psychology, 9*.

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS One, 11*(3), e0150289.

Chouinard, B., Volden, J., Hollinger, J., & Cummine, J. (2019). Spoken metaphor comprehension: Evaluation using the metaphor interference effect. *Discourse Processes, 56*(3), 270–287.

Colangelo, A., Buchanan, L., & Westbury, C. (2004). Deep dyslexia and semantic errors: A test of the failure of inhibition hypothesis using a semantic blocking paradigm. *Brain and Cognition, 54*(3), 232–234.

Colangelo, A., Stephenson, K., Westbury, C., & Buchanan, L. (2003). Word associations in deep dyslexia. *Brain and Cognition, 53*(2), 166–170.

Coltheart, M. (1981). Disorders of reading and their implications for models of normal reading. *Visible Language, 15*(3), 245–286.

Coltheart, M., Patterson, K., & Marshall, J. C. (1987). *Deep dyslexia* (2nd ed.). London: Routledge & Kegan Paul.

Coulson, S. (2008). Metaphor comprehension and the brain. In R. W. GibbsJr. (Ed.). *The Cambridge handbook of metaphor and thought*Cambridge: Cambridge University Press pp. 177–134.

Durda, K., & Buchanan, L. (2008). WINDSORS: Windsor improved norms of distance and similarity of representations of semantics. *Behavior Research Methods, 40*, 705–712.

Gernsbacher, M. A., Keysar, B., Robertson, R. R. W., & Werner, N. K. (2001). The role of suppression and enhancement in understanding metaphors. *Journal of Memory and Language, 45*, 433–450.

Gibbs, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning.* Cambridge University Press.

Ianni, G. R., Cardillo, E. R., McQuire, M., & Chatterjee, A. (2014). Flying under the radar: Figurative language impairments in focal lesion patients. *Frontiers in Human Neuroscience, 8*, 871.

Jacobs, A. M., & Kinder, A. (2018). What makes a metaphor literary? Answers from two computational studies. *Metaphor and Symbol, 33*(2), 85–100.

Jarvis, B. G. (2012). *DirectRT (Version 2012) [Computer software]*. New York, NY: Empirisoft Corporation.

Jones, L. L., & Estes, Z. (2006). Roosters, robins, and alarm clocks: Aptness and conventionality in metaphor comprehension. *Journal of Memory and Language, 55*(1), 18–32.

Katz, A. N., & Al-Azary, H. (2017). Principles that promote bidirectionality in verbal metaphor. *Poetics Today, 38*(1), 35–59.

Kintsch, W. (2000). Metaphor comprehension: A computational theory. *Psychonomic Bulletin & Review, 7*(2), 257–266.

Kintsch, W. (2001). Predication. *Cognitive Science, 25*(2), 173–202.

Kintsch, W. (2008). How the mind computes the meaning of metaphor. In R. W. GibbsJr. (Ed.). *The Cambridge handbook of metaphor and thought* (pp. 129–142). Cambridge: Cambridge University Press.

Kintsch, W., & Bowles, A. R. (2002). Metaphor comprehension: What makes a metaphor difficult to understand? *Metaphor and Symbol, 17*(4), 249–262.

Malhi, S. K., McAuley, T. L., Lansue, B., & Buchanan, L. (2018). Concrete and abstract word processing in deep dyslexia. *Journal of Neurolinguistics*. https://doi.org/10.1016/j.jneuroling.2018.11.001.

Mancopes, R., & Schultz, F. (2008). Processing of metaphors in transcortical motor aphasia. *Dementia & Neuropsychologia, 2*(4), 339–348.

Nenonen, M., Niemi, J., & Laine, M. (2002). Representation and processing of idioms: Evidence from aphasia. *Journal of Neurolinguistics, 15*(1), 43–58.
Newton, P., & Barry, C. (1997). Concreteness effects in word production but not word comprehension in deep dyslexia. *Cognitive Neuropsychology, 14*(4), 481–509.
Ortony, A. (1975). Why metaphors are necessary and not just nice. *Educational Theory, 25*(1), 45–53.
Riley, E. A., & Thompson, C. K. (2010). Semantic typicality effects in acquired dyslexia: Evidence for semantic impairment in deep dyslexia. *Aphasiology, 24*(6–8), 802–813.
Roncero, C., & de Almeida, R. G. (2015). Semantic properties, aptness, familiarity, conventionality, and interpretive diversity scores for 84 metaphors and similes. *Behavior Research Methods, 47*(3), 800–812.
Schmidt, G. L., Kranjec, A., Cardillo, E. R., & Chatterjee, A. (2010). Beyond laterality: A critical assessment of research on the neural basis of metaphor. *Journal of the International Neuropsychological Society, 16*(1), 1–5.
Utsumi, A. (2005). The role of feature emergence in metaphor appreciation. *Metaphor and Symbol, 20*(3), 151–172.