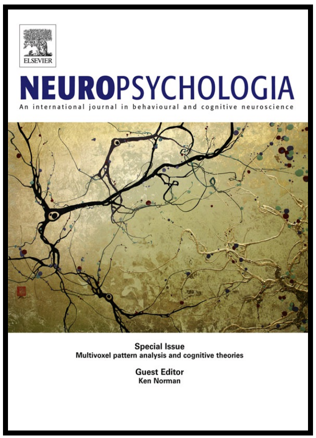
Paper:

Playfoot, D., Billington, J. & Tree, J. (2018). Reading and visual word recognition ability in Semantic Dementia is not predicted by semantic performance. *Neuropsychologia*
http://dx.doi.org/10.1016/j.neuropsychologia.2018.02.011



www.elsevier.com/locate/neuropsychologia

Reading and visual word recognition ability in Semantic Dementia is not predicted by semantic performance

David Playfoot, Jac Billington, Jeremy J. Tree

PII: S0028-3932(18)30069-1
DOI: https://doi.org/10.1016/j.neuropsychologia.2018.02.011
Reference: NSY6684

To appear in: *Neuropsychologia*

Received date: 14 August 2017
Revised date: 27 January 2018
Accepted date: 8 February 2018

Cite this article as: David Playfoot, Jac Billington and Jeremy J. Tree, Reading and visual word recognition ability in Semantic Dementia is not predicted by semantic performance, *Neuropsychologia,* https://doi.org/10.1016/j.neuropsychologia.2018.02.011

Reading and visual word recognition ability in Semantic Dementia is not predicted by semantic performance

David Playfoot[1] Jac Billington[2] and Jeremy J. Tree[3]

[1] Department of Psychology, Sociology and Politics, Sheffield Hallam University, Collegiate Crescent, Sheffield, S10 2BQ, email d.playfoot@shu.ac.uk

[2]School of Psychology, University of Leeds, Leeds LS2 9JT, email j.billington@leeds.ac.uk

[3] Department of Psychology, Swansea University,

Singleton Park, Swansea, SA2 8PP, email j.tree@swansea.ac.uk

Short title: Reading in Semantic Dementia

Corresponding author: David Playfoot, Department of Psychology, Sociology and Politics, Sheffield Hallam University, Collegiate Crescent, Sheffield, S10 2BQ

Email: d.playfoot@shu.ac.uk

Keywords: Semantic Dementia, Surface Dyslexia; Reading; Lexical Decision; Semantics

Abstract

This paper describes longitudinal testing of two Semantic Dementia (SD) cases. It is common for patients with SD to present with deficits in reading aloud irregular words (i.e. surface dyslexia), and in lexical decision. Theorists from the connectionist tradition (e.g. Woollams, et al, 2007) argue that in SD cases with concurrent surface dyslexia, the deterioration of irregular word reading and recognition performance is related to the extent of the deterioration of the semantic system. The Dual Route Cascaded model (DRC; Coltheart et al, 2001) makes no such prediction. We examined this issue using a battery of cognitive tests and two structural scans undertaken at different points in each cases time course. Across both cases, our behavioural testing found little evidence of a key putative link between semantic impairment and the decline of irregular word reading or lexical decision. In addition, our neuroimaging analyses suggested that it may be the emergence of atrophy to key neural regions both inside and outside the anterior temporal lobes that may best capture the emergence of impairments of irregular word reading, and implicated inferior temporal cortex in surface dyslexia.

1. Introduction

Semantic dementia (SD) is a variant of primary progressive aphasia (PPA) in which the core deficit is of semantic memory, a key feature that differentiates it from other PPA variants (Code, Tree & Dawe, 2009; Tree, Kay & Perfect, 2005; Tree & Kay, 2015). As a consequence SD patients score poorly on tasks such as picture naming (Hodges, Graham & Patterson, 1995), word-picture matching (Lambon Ralph & Howard, 2000) and tests of semantic association such as the Pyramids and Palm Trees test (Howard & Patterson, 1992). In contrast, phonological processing in these cases is claimed to be relatively intact (Hodges, Patterson, Oxbury & Funnell, 1992; Snowden, Neary & Mann, 1996) in that performance on such tasks as phoneme discrimination, segmentation and rhyme generation tasks has been reported normal in the face of the dementia (Jefferies, Jones, Bateman & Lambon Ralph, 2005; Knott, Patterson & Hodges, 1997; 2000). In this paper we address the impact that this pattern of performance - semantic deficits coupled with putative intact phonological processing - has on the deterioration of reading aloud over time in semantic dementia. It is common for SD patients to present with surface dyslexia, so much so that surface dyslexia has been considered a cardinal feature (Hodges et al, 1992). Surface dyslexic reading is characterised by a) accurate reading of non-words and words with predictable spelling to sound correspondences, b) poor accuracy for words with unusual spelling to sound patterns and c) a tendency for errors to be *regularisations*, such that "pint" is pronounced to rhyme with "mint." Surface dyslexics will nearly always mispronounce unpredictable (*irregular*) words that are low in frequency, and more severe cases will be in accurate with higher frequency irregular words as well (Bub, Cancelliere & Kertesz, 1985; McCarthy & Warrington, 1986; Shallice, Warrington & McCarthy, 1983). At least in the initial stages of surface dyslexia, non-word reading remains largely unaffected (e.g. Bub et al, 1985; Hodges, Graham & Patterson, 1995).

1.1 The role of semantics in surface dyslexia

Computational connectionist "triangle" models of reading (e.g., Plaut, McClelland, Seidenberg & Patterson, 1996; Seidenberg & McClelland, 1989) make the important prediction that in SD patients who exhibit a surface dyslexic pattern of reading, the extent of the damage to the semantic system is predictive of the severity of the surface dyslexia. Such models (see Figure 1) posit that pronunciation of written words is achieved through the pooled activity of two pathways – one pathway that utilises semantics (Orthographic-Sematic-Phonological: O-S-P) and a second that does not (O-P). The O-P pathway computes pronunciation on the basis of spelling to sound patterns which are learned through exposure to the language. The O-P pathway is especially useful for reading novel words (nonwords) and words for which the spelling to sound mapping is predictable (*regular* words e.g. MINT). For words with less common or less predictable spelling to sound patterns (*irregular* words such as PINT) the O-P pathway is inefficient and inaccurate. As a consequence, the model posits that irregular words can be read via an orthography to semantics to phonology (O-S-P) pathway. Thus, for most people, reading of such items will involve semantic processing to some degree. As we have earlier established, for SD patients the semantic system is compromised and hence it is assumed that the O-S-P route is also disrupted. Thus it follows that the 'triangle' account argues for a relationship between the damage to the semantic system and impaired irregular word reading in surface dyslexia. At the same time, the same model argues that intact phonological processing in SD cases allows non-word reading to remain unaffected by the progression of the dementia (we return to this below).

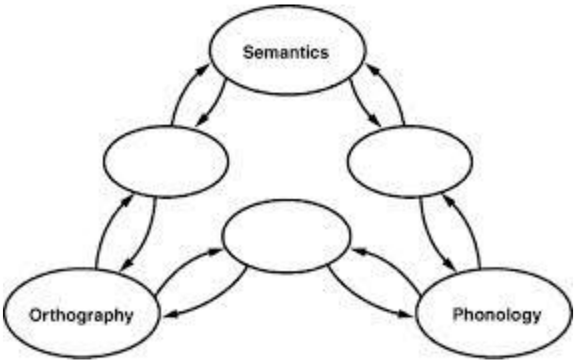


Figure 1 - An example structure of a Triangle model

An alternative account of surface dyslexia has been provided in the framework of the dual route cascaded model of reading (DRC; Coltheart, Perry, Rastle, Langdon & Ziegler, 2001). This model proposes two methods for converting written words into sounds. Reading aloud of non-words is achieved by applying a rule system which associates letters with the most common phonological conversion. This is the grapheme to phoneme correspondence (GPC) route. Pronunciation of irregular words in DRC is achieved by accessing stored orthographic and phonological information from a lexicon containing every word in the reader's vocabulary. Regular words can be read correctly via either of the proposed routes because they conform to the rules prescribed by the GPC route, and are also stored in the lexicon. Importantly, and in contrast to the triangle models discussed above, DRC does not require semantics to be intact for successful irregular word reading, given there is a direct orthography – to – phonological route (see Figure 2). Thus the DRC account predicts that damage to the semantic system need not cause any detriment to reading provided that the rest of the lexical system remains intact and, further, that the extent of the damage to semantic processing is not necessarily tied to the severity of the patient's surface dyslexia.

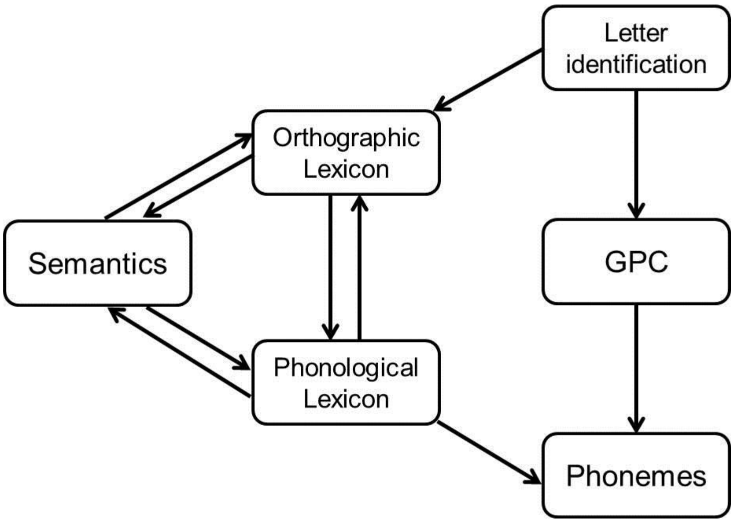


Figure 2 - The dual-route model of reading and word recognition. GPC = Grapheme - Phoneme Conversion system

The key role of semantic information in reading aloud for surface dyslexic cases has been the subject of considerable debate. A particularly important study attempting to address the relationship between semantic performance and irregular word reading was conducted by Woollams, Lambon Ralph, Plaut and Patterson (2007). The paper described a group of 51 Semantic Dementia cases with a view to assessing the prevalence and severity of surface dyslexia. Patients were tested on reading of regular and irregular words (as an index of the severity of their surface dyslexia), and in addition reading performance was compared with a composite score of semantic ability based on word-picture matching and picture naming. As a group, the SD cases included in this analysis exhibited the typical surface dyslexic reading pattern – there was a significant interaction between word frequency and regularity such that irregular word reading was less accurate for low than high frequency words. Woollams et al (2007) also observed a significant relationship between severity of semantic impairment and impairment on word reading. Specifically, patients with lower semantic performance scores were also less accurate in their reading of all classes of words. A relationship was also observed between scores on the composite semantic ability measure and the proportion of regularisation errors (i.e., misreading an irregular word like ‘pint’ to rhyme with mint). A greater proportion of such errors was observed in those cases with lower semantic performance. On the face of this evidence, it appears that the predictions of triangle models were supported – the severity of the semantic deficit is predictive of the severity of surface dyslexia in SD cases. Indeed Woollams et al (2007) successfully simulated this reading pattern in the triangle framework by disrupting the O-S-P reading pathway in their computational model.

However, the data collected by Woollams et al (2007) can also be accommodated by the DRC account. Coltheart, Saunders and Tree (2010) demonstrated that the deletion of a proportion of the lowest frequency words in the model’s orthographic lexicon (the store of visual information about words – see Figure 2) caused irregular word reading errors (i.e. inaccurate irregular word reading is a function of impairment to the lexical route). Regular word reading was preserved, however, as it could be achieved via the intact GPC route. Overall, the more lexical entries were deleted, the greater the use of the remaining intact GPC for all word reading and thus the greater the rate of irregular word regularisation errors. In addition, deleting a proportion of the least used GPC rules created a non-word reading deficit with little effect on regular or irregular word reading. By varying these two proportions of disruption to the routes proposed in DRC, Coltheart et al (2010) were able to closely match data in all of the patients (from the Woollams data set) they simulated without reference to semantics.

There were however some anomalous findings from the Woollams et al’s (2007) study that at first glance appear to be problematic for the triangle account, while also being easily reconciled with the DRC model. In 5 of the 51 cases described, preserved reading was observed *in spite of* poor comprehension. Woollams et al (2007) argued that cases showing such a dissociation between semantic decline and reading performance are rare (although they comprised 20% of their sample), and that when they are observed it is usually soon after the diagnosis of Semantic dementia. In order to account for these data they attributed the findings to individual differences, in that dependence on the semantic system for irregular word reading likely varies between people. It was argued that the *division of labour* between O-P and O-S-P pathways in the reading system is determined while an individual is learning to read, and this may not be uniform across individuals. For any given individual, the degree of reliance on the semantic system for irregular word reading may be contingent on a number of factors relating to the timing and method of instruction in reading acquisition. Thus it is theoretically possible for the semantic system to be completely abolished in an individual who did not pre-morbidly rely on semantics during irregular word reading without leading to surface dyslexia; an individual for whom irregular word reading was heavily contingent on semantics would exhibit surface dyslexia after even mild damage to the semantic system.

On reflection of the data from the Woollams et al (2007) study, it is apparent that patients can vary considerably in the level of their irregular word reading impairment, and thus an individual differences account may have merit. However, retrospective testing of the role of semantics in pre-morbid irregular word reading is clearly impossible, and thus this account is effectively *unfalsifiable* by cross-sectional study. Nonetheless, such a prediction *could* be examined through the longitudinal study of semantic dementia cases. It follows from the connectionist account that SD patients who *do* exhibit surface dyslexia do so because, a) they require semantic information to read irregular words and b) they can no longer access this information. If this is true, then a clear prediction of such an account is that there will be a correlation between semantic performance and irregular word reading accuracy in SD cases with surface dyslexia. Suppose that at a particular point in disease progression two cases present with the same level of semantic difficulties, but that their irregular word reading performance differs – the individual differences account would suggest that the patient with higher irregular word reading accuracy was less reliant on semantics in irregular word reading before the onset of the disease. However, such an account would also make the prediction that the consequences on reading of further semantic decline will differ across these two patients. Simply put, irregular word reading performance would be adversely affected by a smaller decline in semantic performance in a patient who usually relied on semantics to read irregular words. We tested this prediction in this study.

1.2 Non-word reading in acquired dyslexia

A key theme that triangle models embody is the idea that the processes involved in reading (being a relatively recent evolutionary development) are based on pre-existing "primary" systems that serve more general abilities (Patterson & Lambon Ralph, 1999) – in effect reading impairments are never *reading only* impairments. In the previous section, we discussed the notion that dysfunction in primary semantic systems has a detrimental effect on word reading processes. Unsurprisingly, a similar primary systems theme underpins the triangle account of disruption to non-word reading – in this case the key 'primary' system being implicated relates to phonological processing. Evidence for this assertion has been drawn from cases of phonological dyslexia for whom non-word reading is inaccurate but word reading is maintained (see Tree, 2008 for a review). Harm and Seidenberg (2001) demonstrated that a selective impairment of non-word reading could be induced in a triangle model by damaging the phonological attractor network. They argued that print to pronunciation conversion was more difficult when the phonological representations were degraded. Again though, the DRC model differs in its characterisation of the non-word reading deficit in phonological dyslexia – being a reading specific account for impairment, there is no need to assume that nonword reading impairment and generalised phonological impairment must go hand in hand (though such co-occurrence is by no means problematic). For the most part, there have been a number of cases of phonological dyslexia described in the literature who have additional deficits on various phonological tasks (see Tree, 2008 for a review), and this evidence has been argued to support the primary systems account (Harm & Seidenberg, 2001). However, there are also reports of phonological dyslexic cases *without* any generalised phonological impairment (Caccappolo-van Vliet, Miozzo & Stern, 2004a; Caccappolo-van Vliet, Miozzo & Stern, 2004b; Derouesne & Beauvois, 1985; Macoir, Fossard, Saint-Pierre & Auclair-Ouellet, 2012; Tree & Kay, 2006; Tree & Playfoot, 2015). Clearly, such cases constitute a problem for this account, but can be accommodated neatly within the DRC framework which argues they occur as a result of dysfunction to the non-lexical GPC route (Coltheart, 1996). Nickels, Biedermann, Coltheart, Saunders and Tree (2008) demonstrated that the pattern of phonological dyslexic reading (with and without generalised phonological impairment) could be successfully simulated by changing parameters that slow down the function of the GPC route – that is to say, non-word reading impairment could be simulated without damaging the general phonological system.

At present no work has longitudinally examined the relationship between phonological ability and non-word reading in Semantic dementia. As we have already noted, studies have reported that performance on phonological tasks remains within the normal range in cases of SD (Hodges et al, 1992; Jefferies et al, 2005; Knott et al, 1997; 2000; Snowden et al, 1996), and non-word reading is also generally successful, particularly early in the progression of the disorder (Bub et al, 1985). It is of note that of the 34 cases tested on nonword reading in Woollams et al's (2007) study, 24 showed deficits in *both* non-word and irregular word reading, indicating the potential for non-word reading performance to decline over time in this condition. We aimed to further explore this issue, by longitudinally tracking the relationship between nonword reading performance and general phonological ability.

1.3 Longitudinal assessment of reading performance

We consider that it is likely, given that SD is a progressive disorder, that any effect on reading is *also* progressive. Blazely, Coltheart and Casey (2005) discussed how reading ability in SD might deteriorate over time. At the onset of the dementia, reading comprehension is adversely affected but irregular word reading is not. After this initial stage, performance in irregular word reading will begin to deteriorate while regular word reading and non-word reading are preserved. In the final, most severe stage, reading is additionally impaired for non-words (and sometimes regular words too). In sum, stage 1 – no reading impairment, stage 2 – surface dyslexia, stage 3 – generalised reading deficit. Following this account, Coltheart et al (2010) simulated impairment progression within the DRC framework. At the early stages of SD, “pure” surface dyslexia is the result of an impaired lexical route operating alongside an intact GPC route. As the dementia progresses the GPC route becomes damaged as well, eventually resulting in generalised reading impairment. Following up from this work, the present study seeks to provide evidence for these proposals via longitudinally study of the same cases throughout disease progression – as such it is the first to do so in such detail.

1.4 Semantics in word recognition

Thus far we have been discussing reading aloud. Another common task in the assessment of reading processes is the lexical decision task, in which participants are asked to determine whether or not a string of letters presented to them is a real word. Here too the role of semantics is contentious. In terms of the DRC model, it is proposed that every word in the reader's vocabulary is stored as a single unit in the orthographic lexicon (see Figure 2). The presentation of a written stimulus triggers a search of the lexical store. An affirmative lexical decision response ("this is a word") is given if the written stimulus matches an existing orthographic representation. Importantly the lexical decision response can theoretically be achieved on the basis of orthography alone (although semantic information may be beneficial to generating responses quickly, e.g. Balota, Cortese, Sergent-Marshall, Spieler & Yap, 2004). In contrast, triangle models do not propose that there is a store containing individual representations for each known word. Instead, knowledge of the orthography of words is coded as patterns of activation across a large number of units. Discrimination between words and non-words in this framework has only been successfully simulated on the basis of the utilisation of semantic information (e.g. Plaut, 1997; **Plaut & Booth, 2006**). As a consequence, the primary systems approach argues that (just as with irregular word reading), disruption to the primary system of semantic memory has a knock on effect for visual word recognition.

However, there are several reports in the literature, such as those highlighted by Coltheart (2004), of patients exhibiting semantic deficits alongside normal levels of lexical decision accuracy. As such, this evidence is apparently problematic for the triangle model/primary systems account – not for the DRC model, in that it argues for the independence of word recognition and semantics and that the activation of an entry in the lexicon is all that is required and this can be achieved through orthography alone if need be. However, advocates of the primary systems position remained sceptical of reports of problematic cases with apparently ‘normal’ lexical decision performance, citing the utilisation of overly simple test materials or procedures. Rogers, Lambon Ralph, Hodges and Patterson (2004) considered that the impact of a semantic deficit on lexical decision responses could be moderated by orthographic typicality. Specifically, they argued that words that had an unusual spelling or were low in frequency would require a greater semantic contribution before they were recognised, and hence were more susceptible to damage to the semantic system. To examine this under more challenging conditions, Rogers et al (2004) used a two alternative forced choice variation of the standard lexical decision task in which pairs of stimuli were presented simultaneously (one word the other a nonword). Nonword items were also a plausible alternative spelling of the real word (i.e. a pseudohomophone - SKOOL), whilst word items varied in terms of written frequency (high versus low). Critically, one of the items in each pair was higher in orthographic typicality than the other (as measured by bigram and trigram frequency)[1]. Rogers et al (2004) reported an interaction between target frequency and orthographic typicality. When the target word was low in frequency, SD cases were significantly more likely to select a nonword that was high in orthographic typicality. No such effect was observed when word targets were high in frequency. This pattern was exaggerated in cases with more severe semantic impairments, supporting Rogers et al’s prediction.

As before, though, it has been demonstrated that data initially intended to support the triangle model’s predictions can be simulated successfully by the DRC framework. Coltheart, Saunders and Tree (2010) replicated Rogers et al’s (2004) findings by probabilistically deleting low bigram frequency words from the orthographic lexicon of the model. These

---

[1] While Rogers et al (2004) describe their stimuli as being orthographically typical on the basis of bigram and trigram frequencies, it is also the case that they are more word-like in relation to their orthographic neighbourhood. The "typical" items in the stimulus sets had a higher N (the number of real words that could be made by changing a single letter; Coltheart, Davelaar, Jonasson, & Besner, 1977) than the atypical sets.

lesions were sufficient to mirror the patterns in Rogers et al's (2004) patients without the need for disruption to the semantic system. In effect, the simulation work demonstrated that, under such circumstances, representations for high bigram frequency words remained accessible (and thus no effect of orthographic typicality in lexical decision), but for low bigram frequency words it was possible that the entry may have been deleted from the lexicon (and thus lexical decision performance would be impaired). Thus, in relation to word recognition, as in reading aloud, the importance of semantics is the subject of considerable debate across the two theoretical accounts. Again, we contend that longitudinal study of cases of SD may provide some clarity. If the function of the semantic system is the key to lexical decision performance then word recognition accuracy should be related to scores on tests of semantic processing in a predictable manner – as semantic processing fails so should lexical decision. On the other hand, if word recognition can be achieved without needing to refer to semantics (as in the DRC account) then lexical decision accuracy may be independent of semantic performance.

1.5 Neural bases of reading

To foreshadow the work relating to our specific study, we sought to undertake some longitudinal neuroimaging work (via two MRI scans over time) with our two semantic dementia cases, in order to consider the pattern of neurodegenerative changes relative to the longitudinal changes seen in our behavioural testing. As a consequence, it is worth considering the literature on semantic dementia and neuroimaging work that implicates key neural regions in the processes of irregular word reading and word recognition. Our goal in this section is not to dissect the neuroimaging literature in any great detail as a number of excellent reviews already exist. Here we simply wish to highlight the studies that are particularly pertinent to the debate about the role of semantics in reading processes in Semantic Dementia and surface dyslexia. The cognitive theories outlined earlier in this introduction suggest that different processes are required to read different types of written stimulus. Henson (2005; 2006) suggested that the *functional* separation of these processes could be mapped onto distinct brain areas. This assumption is the basis for the analysis of neuroimaging data in the literature, and in the current study.

Theorists from the triangle perspective suggest that word recognition and irregular word reading are tasks that require the semantic system, and hence they ought to implicate brain regions where semantic processing takes place and which deteriorate in semantic dementia. A key region for semantic processing considered in this study is the bilateral anterior temporal lobes (ATLs) as a likely neural proxy. Converging evidence has suggested that these regions are important for semantic memory. Pertinent to this work is the fact that patients with SD typically show atrophy of the ATL, and that the degree of neural degeneration is correlated with performance on semantic processing tasks (e.g., Davies, Graham, Xuereb, Williams & Hodges, 2004; Noppeney et al, 2007). In addition, evidence from functional magnetic resonance imaging (fMRI) studies of healthy participants when completing semantic tasks also implicates the ATL. In a meta-analysis of 164 such studies, Visser, Jefferies and Lambon Ralph (2009) demonstrated that the ATLs were activated during semantic processing (though both the precise coordinates of the activity and the strength of the activity was influenced by a number of methodological variables). As a consequence, this study will assume damage to the ATL regions in semantic dementia is indeed a reasonably neural proxy for disruption to the semantic system, and thus under the predictions of the triangle perspective, irregular word reading and lexical decision performance should be worse when the ATL is more severely atrophied (since progressive semantic deterioration and poorer performance on word reading/lexical decision should go hand in hand). However, SD is by definition a neurodegenerative disease, and as such, patients will show an overall atrophic profile that implicates a great many other neural regions than just the ATL. It has been suggested that perhaps neural regions *outside* of this key area may play a role in the emergence of surface dyslexia in such patients. We are not the first authors to discuss these possibilities. While they acknowledged that it was difficult to pinpoint the precise location through lesion studies, McCarthy and Warrington (1990) suggested that surface dyslexia was linked to damage in the posterior middle temporal areas of the left hemisphere. As the atrophy in semantic dementia cases progresses to regions posterior and superior to the ATL (the temporal occipital cortex and the inferior frontal cortex, respectively) over time, it has been argued that damage which initially affects the semantic system spreads to brain regions involved in the mapping from orthography and phonology. The resulting detriment to reading aloud may be mistakenly attributed to a worsening semantic deficit purely because the semantic deficit was evident before the surface dyslexia. Furthermore, patients in whom these posterior and superior regions have remained relatively spared have retained some language functioning despite severe decline in semantic processing (e.g. patient BS, Bright, Moss, Stamatakis, & Tyler, 2008, see also Blazely et al, 2005).

Other work undertaken by Brambati et al, (2009) found that exception word accuracy in SD was linked to with gray matter volume in the left anterior temporal structures, *including* the temporal pole, the anterior superior and middle temporal and fusiform gyri. Thus, on reflection, it is apparent that the neurodegenerative profile of SD can include atrophy that extends throughout the temporal lobe (both anterior to posterior and inferior to superior), and as a consequence it is far from clear whether the critical region for atrophy is indeed the ATL (which as we have suggested is a proxy for semantic memory loss), but rather may implicate other neural regions.

In recent work outside of the field of SD, Taylor, Rastle and Davis (2013) performed a meta-analysis of fMRI studies of reading in healthy adults, contrasting areas that were activated during reading of words and pseudowords. Their analysis identified a number of areas differentially implicated in reading the two types of stimuli, and Taylor et al (2013) mapped these areas onto theoretical components to find the "best fit" between neural data and cognitive models. Areas in the posterior temporal lobe (pTL) and fusiform gyrus (FG) were considered to be a good candidate for the orthographic system (in triangle models) or the letter units, extending into the orthographic lexicon (in DRC) – and this is consistent with our point above that atrophy to the posterior temporal lobe may be key to the emergence of surface dyslexia. Indeed, other recent imaging work with a different neuropsychological population has implicated these posterior temporal regions specifically in the processing of irregular words during reading aloud. Binder, Pillay, Humphries, Gross, Graves and Book (2016) studied 45 patients who had suffered stroke, and had subsequently presented with surface dyslexia. **The patients were presented with a list of words that varied frequency and regularity, and a test of semantic association very similar to the Pyramids and Palm Trees test (Howard & Patterson, 1992, described in more detail in the method section of the current paper). The behavioural pattern described by Binder et al (2016) was typical of surface dyslexia - a greater deficit in reading aloud irregular versus regular words, particularly those which were low in frequency, and a high proportion of regularisation errors. Notably, there was no correlation between the scores on the semantic association tasks and the number of regularisation errors (i.e. semantic performance was not related to irregular word reading performance). Of greatest importance for the current study, though, are the findings of the voxel-based lesion-**

symptom mapping analyses Binder et al (2016) reported. They identified the lesion sites that were associated with a) the commission of regularisation errors in reading aloud and b) failure to identify semantic associations correctly. The authors reported a strong positive relationship between the extent of the damage to the left posterior middle temporal gyrus and the prevalence of regularisation errors among these patients that did not overlap with the regions of damage that were related to errors in the association task. Irregular word reading was compromised by lesions in areas that did not also cause semantic processing deficits. Binder et al (2016) suggested that similar surface dyslexic reading patterns might arise from two different lesion loci. On the one hand, they suggested that damage to the ATL may lead to surface dyslexia *and* semantic impairment. On the other, they suggested that damage to more posterior areas of the temporal lobe may lead to surface dyslexia *without* semantic impairment.

In summary, it would appear that a number of studies with SD (as well as stroke patients and healthy adults) suggests that the relationship between irregular word reading and key neural regions is by no means limited to what might be occurring in the ATL - and thus it would be interesting to explore the pattern of changes across the temporal lobe longitudinally in our SD patients. This may speak to the above debate regarding which neural regions (and their emerging atrophy) may be key to the presence and severity of surface dyslexic impairment for both our cases.

1.6 The current study

We assessed two Semantic Dementia cases with regard to their performance in reading words and non-words, their semantic ability and their phonological skill. Specifically, we attempted to address five questions. First, we aimed to assess whether reading deficits associated with

SD evolve in the way suggested by Blazely et al (2005). Second, we assessed the relationship between semantic performance and irregular word reading over time in order to test two different theoretical accounts. To reiterate, the primary systems account suggests that the severity of irregular word reading deficits will be contingent on the level of semantic decline; the DRC account does not posit a predictable relationship between these abilities. Third, we examined whether non-word reading accuracy also declines over time, and if so, what relationship this has with phonological performance. Again, the primary systems account argues for a relationship between disruption of general phonological processes and impaired non-word reading, whereas DRC makes no such prediction. Fourth, we examined whether there is a predictable relationship between semantic performance and lexical decision accuracy. Here, too, a triangle model embodying a primary systems account predicts that successful word recognition is contingent on an intact semantic system, while DRC allows for word recognition to occur independently of semantics. Finally, we explored the structural changes in grey matter volume across the brain in both our cases over disease progression using VBM (Ashburner, 2000), in order to determine whether this could shed light on the issue of where along the temporal lobe atrophic change may be key to the emergence of impaired irregular word reading for both our cases.

2. Method

2.1 Patient descriptions

2.1.1 Patient JD

JD was 59 at the time of the first test session. A more detailed description of her case can be found in Playfoot, Izura and Tree (2014). Her performance in phonological and syntactic tasks (Psycholinguistic Assessments of Language Processing in Aphasia, Kay, Lesser & Coltheart, 1992; Test for the Reception of Grammar, Bishop, 1989) was largely normal but she showed considerable deficits on semantic tasks (e.g. Pyramids and Palm Trees, Howard & Patterson, 1992). A summary of JD's scores on standard neuropsychological tests is presented in Table 1. JD showed the classic surface dyslexic reading pattern – accurate with non-words, regular words and words high in frequency, but impaired when pronouncing low frequency irregular words. JD also made errors in lexical decision, all of which were false positives in non-word trials.

2.1.2 Patient NJ

NJ was 56 when he was first tested on the battery used in this study. His semantic processing was impaired at the time of initial testing, but his phonological performance was close to ceiling (see Table 1). NJ exhibited poor irregular low frequency word reading, but reading for high frequency irregular words and regular words (irrespective of frequency) remained flawless. Non-word reading accuracy was also good. NJ did not make any errors in lexical decision at baseline.

2.2 Materials and procedure

We administered the tests in Table 1 above to each case at intervals of between 6 months and one year. Two issues are necessary to note at this point. First, we worked with each case for a different length of time (as JD chose to withdraw), so the maximum number of data collection points was fewer for JD than for NJ. Secondly, not all the tests were administered at every encounter. Our analyses here will only include the number of consecutive test sessions in which the tests were performed, and data points are not more than one year apart.

Table 1 - Baseline neuropsychological test performance (%). Norms are taken from the manuals for the test unless otherwise noted. Impaired performance has been marked with an asterisk.

		JD	NJ	Norms
Semantics	Pyramids & Palm Trees (pictures)	91*	73*	96
	Pyramids & Palm Trees (written)	83*	81*	96
	TROG	98	100	
Reading	PALPA 35 Regular	100	100	100
	PALPA 35 Irregular	87*	87*	100
	PALPA 36 Nonwords	96	100	95
	Weekes HF words	100	100	100
	Weekes LF words	95*	100	100
	Weekes Nonwords	95*	99	100
	Woollams HF Regular words	100	100	100
	Woollams HF Irregular words	95*	100	100
	Woollams LF Regular words	93*	100	100
	Woollams LF Irregular words	74*	88*	98
	Woollams Non-words	98	98	96
Phonological	PALPA 16 Segmentation	100	100	
	PALPA 15 Auditory Rhyme judgement	93	100	93[a]
	PALPA 9 Word repetition	100	100	99
	PALPA 9 Nonword repetition	100	100	94
	Digit span Forward/Back	9/6	9/7	-
Lexical decision	PALPA 25 Imageability x Frequency	88*	100	99
	PALPA 27 Regular	100	100	100
	PALPA 27 Irregular	100	100	99
	PALPA 27 Nonwords	67*	100	99
	PALPA 27 Pseudohomophones	67*	100	97
	Rogers HF W>NW	83*	100	99
	Rogers HF NW>W	94*	100	100
	Rogers LF W>NW	89*	100	98
	Rogers LF NW>W	72*	100	97

Note: Pyramids and Palm Trees (Howard & Patterson, 1992), TROG = Test for Reception of Grammar (Bishop, 1989), PALPA = Psycholinguistic Assessments of Language Processing in Aphasia (Kay et al, 1992). High and low frequency words and non-words were taken from Weekes (1997). HF = high frequency; LF = low frequency. Norms for the Woollams et al (2007) reading task and the Rogers et al (2004) lexical decision task are from control participants described in section 3.1.

2.2.1 Semantic performance

We used two measures of semantic processing ability: the Pyramids and Palm Trees test (Howard & Patterson, 1992) and the Test for the Reception of Grammar (or TROG, Bishop, 1989). The Pyramids and Palm Trees task is a test of semantic association. Participants are presented with a cue picture. Below the cue are two other pictures and participants are asked to decide which of these is the best match for the cue. There is also a version which replaces the pictures with written words, but otherwise follows the same procedure. The TROG assesses the ability of a participant to process syntax. They are asked to choose which of four pictures matches the meaning of a spoken sentence.

2.2.2 Phonological performance

The phonological ability of our patients was examined using a number of common tasks drawn largely from the Psycholinguistic Assessments of Language Processing in Aphasia battery (PALPA, Kay et al, 1992). We tested rhyme judgement. This task asks the participant to compute the pronunciation of pairs of written words and decide whether they **rhyme** when pronounced. We also assessed segmentation ability using subtest 16 of the PALPA battery. Further, and following the work of Tree and Kay (2006) and Tree (2008), we considered aspects of auditory verbal short term memory on the basis that such skills may be implicated in the non-word reading dysfunction in phonological dyslexia, and may thus inform any eventual non-word reading deficit in SD as well. Patients were therefore tested for digit span and both word and non-word repetition from the PALPA battery (PALPA 9).

2.2.3 Reading

We used four different reading tasks for the purpose of this investigation. First (and most importantly), we presented JD and NJ with the same stimuli used by Woollams et al (2007). This consisted of 168 words varying in frequency (high versus low) and regularity (regular versus irregular), and 40 non-words. We also presented tests of non-word (subtest 36), regular and irregular word (subtest 35) reading from the PALPA battery. Finally we used high frequency words, low frequency words and non-words taken from Weekes (1997).

2.2.4 Lexical decision

In order to test word recognition capabilities we presented three lexical decision tasks from the PALPA battery (subtests 25, 26 and 27). These examined the effects of imageability and frequency, morphology, and spelling to sound regularity respectively. Crucially we also asked our participants to complete the lexical decision task used by Rogers et al (2004).

2.2.5 Voxel Based Morphology Data Analysis

Imaging Data was collected on a 1.5T Phillips MRI scanner. All VBM data analysis was carried out using the CAT12 toolbox (Gaser and Dahnke, 2016) for Statistical Parametric Mapping 12 (SPM, Wellcome Department of Cognitive Neurology, London) running on Matlab R2011a (MathWorks, Natick, MA). For both patients and controls a T1 image was collected using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) with a TR =25ms, TE=4.1288, FA=30deg, voxel size 0.898, 0.898, 0.8 (slice thickness). Structural images were pre-processed using the standard CAT12 longitudinal data pipeline. Each of the participants' longitudinal images (time=1-n) were aligned to AC-PC space and a mean reference image (RI) created. Each image (1-n) was then realigned to the RI and then corrected for any signal inhomogeneities in comparison to the RI. SPM12 probability maps were used in order to segment all images into grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) and GM, WM, and CSF images 1-n were realigned a final time.

An adapted Diffeomorphic Anatomic Registration Through Exponentiated Lie Algebra (DARTEL) algorithm (Yotter, Zieglar, Thompson, and Gaser, 2011d; Ashburner, 2007) was used in order to warp the segmented GM, WM and CSF RI into Normalised (Montreal Neurological Institute (MNI)) space. In order to correct for post-warping volume changes in each tissue, the normalised images were modulated using DARTEL Jacobian determinant maps. Normalised images were smoothed to the full-width Gaussian kernel of 10mm. Statistical analysis (t-tests) were performed in SPM12. A global normalisation correction was applied to each analysis in order to consider GM volumes relative total intracranial volume (TIV=GM+WM+CSF). This procedure corrects for brain size differences across participants. Second level contrasts were run at a cluster Family Wise Error (FWE) corrected threshold of p<0.5 (Friston, Holmes, Poline Price, & Frith, 1996) using an inclusive GM mask and clusters are reported in (Montreal Neurological Institute) MNI space.

3. Results and discussion

The percentage of correct responses offered by JD and NJ in each test and at each test time are presented in Table 2. The baseline assessments are also included as a comparison. In the interests of clarity, each phase of the analysis will be presented in a separate section and the relevant test scores will be reiterated. However, Table 2 shows that overall a) semantic performance declined for both cases, b) phonological performance remained high for both

Table 2 – Percentage accuracy for JD and NJ in tests of semantic, phonological, reading and lexical decision over time. Impaired performance is marked with an asterisk.

		JD					NJ				Norms
		1	2	3	4	5	1	2	3	4	
Semantic	Pyramids & Palm Trees (pictures)	96*	83*	81*	69*	62*	73*	69*	58*	52*	96
	Pyramids & Palm Trees (written)	83*	77*	56*	50*	52*	81*	63*	---	---	96
	TROG	98	94	---	89	65	100	98	---	---	
Reading	PALPA 35 Regular	100	97*	100	83*	---	100	100	100	100	100
	PALPA 35 Irregular	87*	77*	53*	40*	---	87*	90*	93*	90*	100
	PALPA 36 Nonwords	96	96	96	100	96	100	100	100	96	95
	Weekes HF words	100	100	97*	97*	99*	100	98*	100	---	100
	Weekes LF words	95*	95*	90*	93*	86*	100	100	98*	---	100
	Weekes Nonwords	95*	98	95*	95*	---	99	97*	93*	---	100
	Woollams HF Regular words	100	100	100	100	98*	100	100	95	90*	100
	Woollams HF Irregular words	95*	91*	91*	86*	71*	100	93	88	55*	100
	Woollams LF Regular words	93*	83*	90*	88*	88*	100	100	95*	79*	100
	Woollams LF Irregular words	74*	60*	69*	48*	50*	88*	64*	62*	12*	98
	Woollams Non-words	98	93	98	93	98	98	93	90*	83*	96
Phonological	PALPA 16 Segmentation	100	100	100	100	100	100	100	---	---	
	PALPA 15 Auditory Rhyme judgement	93	95	97	95	82*	100	100	---	---	93
	PALPA 9 Word repetition	100	100	100	100	100	100	100	100	100	99
	PALPA 9 Nonword repetition	100	100	100	100	85	100	100	100	100	94
	Digit span Forward/Back	9/6	9/6	8/5	7/4	6/4	9/7	8/6	7/5	6/4	
Lexical decision	PALPA 25 Imageability x Frequency	88*	84*	81*	69*	55*	100	98	98	94*	99
	PALPA 27 Regular	100	87*	93*	100	87*	100	93*	100	93*	100
	PALPA 27 Irregular	100	80*	93*	100	100	100	100	93*	93*	99
	PALPA 27 Nonwords	67*	73*	73*	53*	47*	100	100	100	100	99
	PALPA 27 Pseudohomophones	67*	73*	80*	53*	40*	100	93	100	87*	97
	Rogers HF W>NW	83*	78*	83*	78*	83*	100	83*	78*	83*	99
	Rogers HF NW>W	94*	72*	72*	44*	44*	100	100	100	89*	100
	Rogers LF W>NW	89*	89*	83*	83*	83*	100	89*	89*	67*	98
	Rogers LF NW>W	72*	61*	67*	61*	11*	100	89	78*	67*	97

Note: The segmentation and auditory judgement tasks were not performed by NJ in sessions 3 and 4 because he was not able to understand the instructions for the task. Norms are the same as in Table 1.

cases, c) irregular word reading became less accurate over time, d) non-word and regular word reading only declined in later sessions and e) lexical decision accuracy decreased for both cases. Our inferential analyses are split into three parts. We will first address the question of when (or if) JD and NJ performed significantly worse than normal for tests of semantics, phonology, reading and lexical decision. The second part of the analysis will assess any decline in performance for each of our cases over time. Finally, we will offer comparisons between JD and NJ and the relationships between semantic performance and reading accuracy, semantic performance and lexical decision accuracy, and phonological performance and non-word reading accuracy.

3.1 Comparing JD and NJ with normal performance

In all the analyses reported in this section, Crawford's t-test (Crawford & Howell, 1998) was used to compare the score obtained by the patients in this study with the distribution of scores in a normative population where possible. These norms were taken from the published data accompanying the tests unless otherwise stated.

As both JD and NJ had been diagnosed with SD, it is unsurprising that their performance on semantic tasks was already impaired in the initial session. In relation to phonological tasks, the pattern was also clear cut - JD and NJ performed at or near ceiling throughout the study in word repetition, non-word repetition, segmentation and auditory rhyme judgement. Even when JD's accuracy in non-word repetition did drop slightly in the last test session, the decline did not take her below the normal range [$t(1) = 1.154$, $p > .1$]. JD remained accurate in more than 93% of rhyme judgments until her final session. Neither of our cases made any errors in segmentation in any test session. It appears, therefore, that phonological performance remains unaffected by semantic dementia long after the semantic system becomes degraded.

Next we determined whether significant differences in reading accuracy could be observed between a group of control participants without reading deficits (N=23, mean age = 65, SD = 7.6) and JD or NJ respectively, with low frequency regular and irregular words and non-words[2]. JD already scored significantly lower than controls at baseline for low frequency regular [t(1) = 8.160, p < .001] and low frequency irregular words [t(1) = 10.108, p < .001]. Significant differences in reading low frequency irregular words were also observed between NJ and controls at baseline [t(1) = 4.145, p < .001]. His performance with low frequency regular words was within the normal range until the third test session by which point his accuracy was significantly lower than controls [t(1) = 5.317, p < .001]. With regard to non-word reading, both cases remained highly accurate long after low frequency word reading had declined; JD never performed worse than control participants on this task, and NJ did not show any deficit until the penultimate session [t(1) = 2.400, p < .05].

Finally, analysis of the performance of our patients on the two alternative forced choice lexical decision task conducted by Rogers et al (2004) was conducted. As above, the analyses examined if and when the semantic dementia cases' performance became significantly poorer than healthy controls (N=27, mean age = 64.3 years, SD of age = 6.71) by using Crawford's t-tests. JD's accuracy in all conditions was already significantly below that of our control sample at initial testing [high frequency W>NW t(1) = 7.731, p < .001; high frequency NW>W t(1) = 4.910, p < .001; low frequency W>NW t(1) = 2.455, p < .05; low frequency NW>W t(1) = 3.999, p < .001]. By the second encounter, NJ's accuracy was

---

[2] Crawford's t-test compares a single score to a distribution of scores from a normative population to determine the probability that the new observation is drawn from the same population. In instances where the normative population has a standard deviation of zero (as was the case for high frequency word reading) the test cannot be applied.

significantly below the normal range for two out of four conditions [high frequency W>NW t(1) = 7.731, p < .001; low frequency W>NW t(1) = 2.455, p < .05] but not for low frequency NW>W trials [t(1) = 1.278, p > .1]. His performance for low frequency NW>W trials had decreased below the normal range by test 3 [t(1) = 3.092, p < .01]. NJ remained at ceiling in the high frequency NW>W condition until the fourth test session, at which point his accuracy was significantly impaired [t(1) = 4.910, p < .001].

The analysis of the performance of our cases indicated that a) semantic ability was impaired prior to the first session for both cases and b) phonological ability was never impaired for either case. At the time of initial testing, JD was significantly worse than controls for reading aloud low frequency regular and irregular words, and performed below the normal range in lexical decision too. Her non-word reading, though, was good and remained stable throughout. In contrast, NJ only performed below the normal range in reading aloud low frequency irregular words when we first tested him. His performance declined such that he became less accurate than controls first for lexical decision, then for low frequency regular word reading and finally for non-word reading.

3.2 Progression of performance over time

The next set of analyses looked for significant decreases in semantic ability, reading aloud and lexical decision across test sessions. As we observed that phonological ability did not ever decrease below the normal range for either of our cases, we did not include these tests in our analyses. We compared pairs of scores for each patient using McNemar's test.

3.2.1 Semantic performance

JD's accuracy (Figure 3) on the pyramids and palm trees picture task decreased significantly between the first and second session [X^2 (1) = 5.1429, p < .05]. There were no further significant decreases in performance for consecutive sessions (all p > .1), but JD was significantly less accurate in the last session than she had been in the first [X^2 (1) = 14.45, p < .001]. Overall her accuracy had dropped from 96% to only 62%. For NJ, the only significant decrease in accuracy between consecutive sessions was observed between the penultimate and final sessions [X^2 (1) = 4.5, p < .05], although his overall decrease from 73% in the first session to 52% in the last was significant [X^2 (1) = 12, p < .001]. In the written version of the task, JD's accuracy declined significantly between sessions 2 and 3 [X^2 (1) = 4.7619, p < .05], and her overall decrease from baseline to final testing was significant [X^2 (1) = 18.05, p < .001]. NJ's overall drop from 81% to 63% in the written Pyramids and Palm Trees test approached significance [X^2 (1) = 3.368, p = .066].

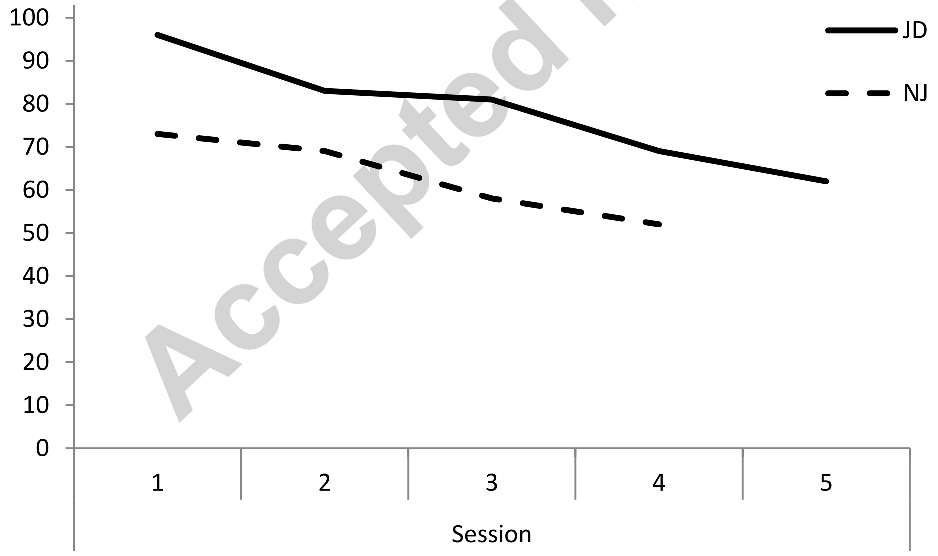


Figure 3 - Percentage correct responses for JD and NJ on the picture version of the Pyramids and Palm Trees (Howard & Patterson, 1992). Note that NJ provided data on only 4 occasions. The absence of a data point does not indicate a score of zero.

In summary, semantic ability declined for both of our cases. NJ's scores dropped rapidly between session 2 and session 3, and then levelled out at near chance. His overall performance in the written version of the task did not decline significantly in the data collection period. For JD there appears to be two different patterns across modalities. Her performance in the written version of Pyramids and Palm Trees plummets at session 3 (dropping from 77 to 56) where it effectively reaches chance levels. For the pictures version, a large decrease happens at session 4 (from 81 down to 69), and then performance stabilises for the remaining testing. It is also of interest to note that JD's overall decline in semantic scores (34% in Pyramids and Palm Trees) was considerably larger than NJ's (21%).

The next question we aimed to address in this paper was as to whether there were stages of surface dyslexia as suggested by Blazely et al (2005). A series of McNemar's tests were performed to determine where significant changes in accuracy were exhibited across consecutive testing phases. The analyses below address reading accuracy for regular words and irregular words separately.

3.2.2 Progression of reading performance

3.2.2.1 Regular words

In relation to high frequency regular words (top left, Figure 4), it is apparent that JD remains well able to read this set of items even late in disease progression, whilst for NJ there is a steady, albeit mild, decline. Neither case decreased significantly across test sessions for these items (all p > .1). For low frequency regular words (bottom left, Figure 4) JD showed a slight decline on these items relative to controls, but this remains relatively stable throughout the testing. By the final session, she is slightly (though not significantly) worse at reading low frequency regular words than at baseline [X^2 (1) = 0.571, p > .1]. NJ, on the other hand, is initially much better with these items, but by the later sessions his performance declines to mirror the level of impairment seen in JD. By the final session, he is considerably poorer at reading low frequency regular words than at baseline [X^2 (1) = 7.111, p < .01].

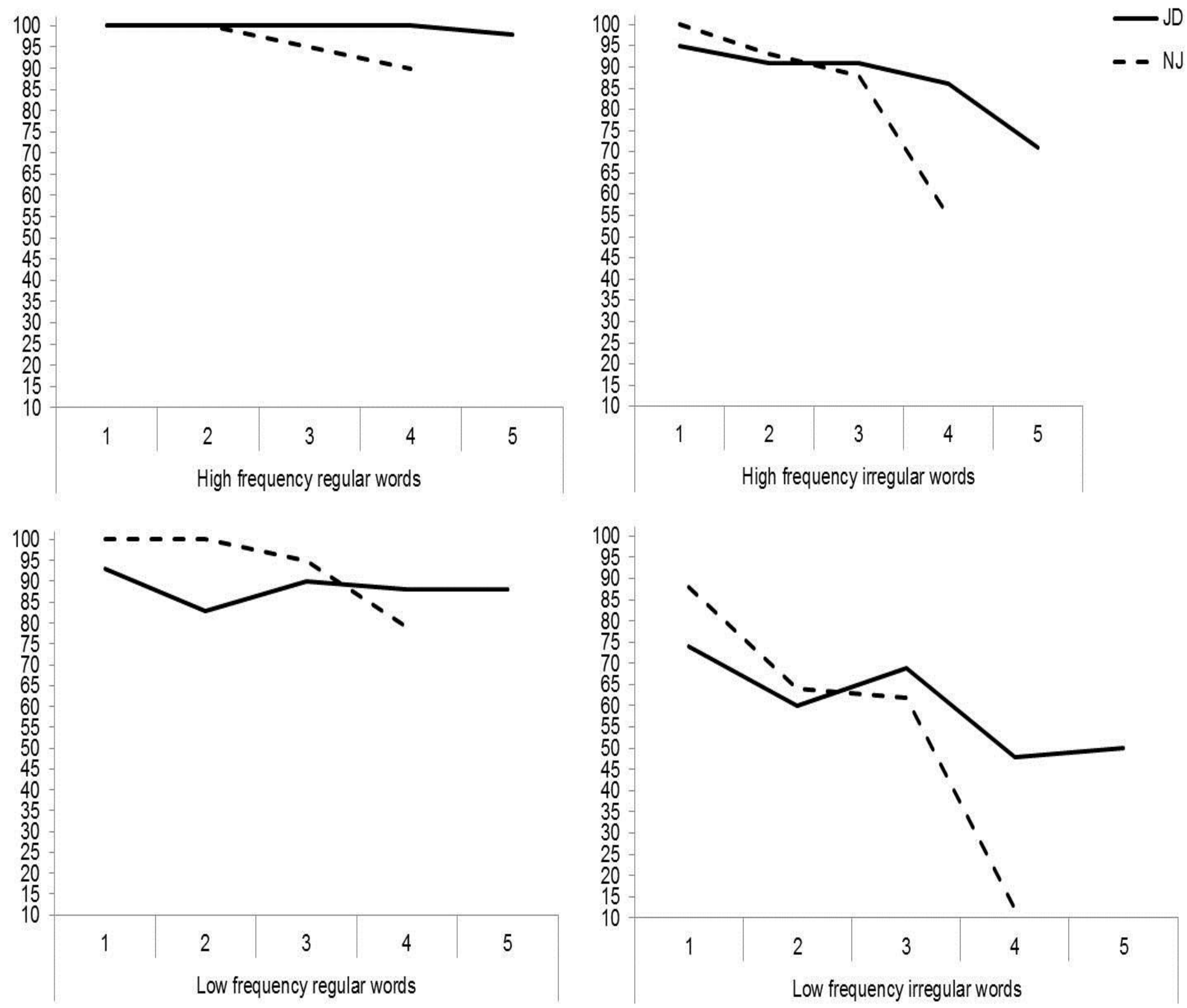


Figure 4 - Percentage correct for each case in each test session. The figure is split into separate panels for each stimulus type. NJ only performed the reading task on 4 occasions.

The pattern of decline observed in our cases suggests that regular word reading, particularly for low frequency words, *does not* remain unaffected by disease progression in semantic dementia. It also appears that this decline is not uniform – for NJ there is a steady decline for low frequency regular words, whilst for JD there is early stage impairment that does not progress further.

3.2.2.2 Irregular words

At initial testing, neither case was substantially impaired with high frequency irregular words (top right, Figure 4). Over time, JD shows a slow decline in performance on these items, with obvious impairment by the final session [$X^2 (1) = 6.75$, $p < .01$ versus baseline]. For NJ there is a sharp decline following the penultimate testing session [$X^2 (1) = 5.882$, $p < .05$] - essentially an accelerated pattern versus that of JD. In relation to low frequency irregular words (bottom right, Figure 4) NJ was only mildly impaired with these items at baseline, whilst JD showed obvious impairment. Over time, JD’s impairment gradually worsens such that her accuracy lower by a significant 24% [$X^2 (1) = 4.9$, $p < .05$]. NJ’s initial mild impairment accelerated very rapidly to near zero performance by final stage. Thus, as with high frequency irregular words, NJ showed a far sharper decline in performance than did JD.

As with our testing on regular word reading, the evidence suggests that the pattern of decline for irregular words is by no means uniform across cases of semantic dementia - for JD there is evidence of a severe impairment with low frequency irregular words at initial testing, and over time this declines further; whilst a similar steady decline is seen with high frequency items. This suggests a steady decline across all forms of irregular word items. For NJ the pattern is a different, in that high frequency irregular word reading remains relatively intact until he exhibited a much accelerated decline, resulting in a marked decrease in performance. It is possible JD might also have shown a sharp decline at later stages, if she had remained willing to be tested – it therefore remains speculative as to whether accelerated decline (seen in NJ) is in fact a typical feature for semantic dementia disease progression at some stage. **The vast majority of errors made in irregular word trials were regularisations for both cases.**

3.2.2.3 The relationship between semantic performance and reading

One of the main aims of this paper was to determine whether a systematic relationship existed between accuracy in reading irregular words and performance on semantic tasks. The DRC account (Coltheart et al, 2001) does not predict a close relationship between semantics and reading. The Triangle account (e.g. Plaut et al, 1996) argues that the two are inextricably linked such that more severe irregular word reading deficits will be observed in instances where the semantic deficit is also severe. A further, and related, prediction made by Woollams et al (2007) is that in cases where semantic performance is not strongly implicated in irregular word reading, it is because there are individual differences in the way in which reading is acquired that result in the role of semantics in normal (pre-morbid) reading being of varied importance. Following this logic, a case for whom irregular word reading remains good in spite of noticeable semantic damage should be less affected by further damage to the semantic system than a case with poor irregular word reading performance. We examined these predictions by comparing the performance of our two cases.

Figure 5 plots the accuracy of each case when reading low frequency irregular words (those items that are most affected in surface dyslexia) against their score on picture version of the Pyramids and Palm Trees in the same session. The comparison of the patients' performance using chi square revealed an interesting pattern. JD scored significantly higher on the Pyramids and Palm Trees test of semantic ability in the first session [X^2 (1) = 18.477, p < .001], but NJ performed significantly better than JD when reading low frequency irregular words [X^2 (1) = 5.491, p < .05]. This pattern can be explained by either DRC or Triangle accounts. However, the Triangle account would have to explain this dissociation by suggesting that NJ was less reliant on semantics for reading irregular words prior to the onset of the semantic dementia. Thus the decrease in his accuracy for low frequency irregular word reading should be smaller than for JD following an equivalent decrease in semantic performance. As we have seen in the previous section, there is evidence that runs counter to this prediction in that NJ showed a marked an accelerated decline in reading of irregular word items by later stages of testing.

In the second and third sessions, again, JD had better scores on the Pyramids and Palm Trees task [session 2 X^2 (1) = 4.633, p < .05; session 3 X^2 (1) = 11.416, p < .01] but this time JD and NJ had statistically similar levels low frequency irregular word reading accuracy (both p > .1). In the fourth session, the last for NJ, the two cases were again significantly different in their score on Pyramids and Palm Trees [X^2 (1) = 5.356, p < .05] with JD performing the better. Now, though, JD was significantly better than NJ at irregular word reading [X^2 (1) = 29.167, p < .001] – consistent with his accelerated decline profile

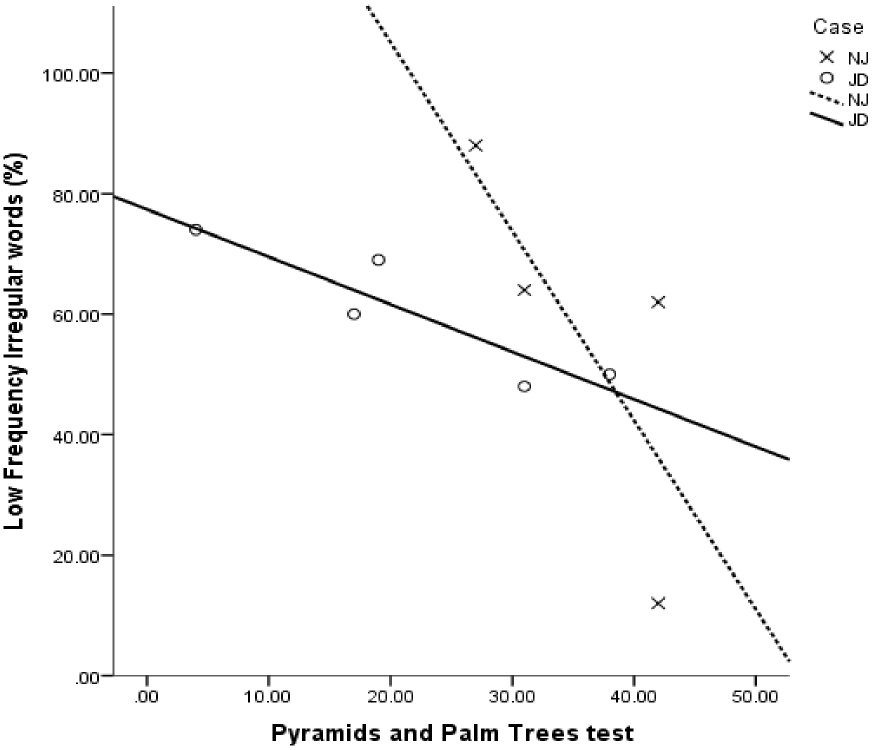


Figure 5 - Percentage correct in reading aloud low frequency irregular words plotted against percentage errors on the Pyramids and Palm Trees test for each case. Scores were provided in different test sessions over time, with an increasingly greater number of errors in the Pyramids and Palm Trees test as the disease progressed. Therefore data points towards the right of the x axis are from later test sessions.

Across these 4 sessions, JD's score on Pyramids and Palm Trees decreased by 33% and NJ's by 21% - suggesting a more marked decline over time of semantics for JD relative to NJ. However, as has been discussed earlier, it was NJ that showed the much more marked and accelerated decline in irregular word reading. This, alongside the regression lines in Figure 5 suggests that the impact of a decrease in semantic performance was greater for NJ than for JD. At this point, it is worth mentioning the recent work by Binney et al (2016). They described the reading performance of a group of semantic dementia cases (particularly the regularity effect) and used both the word and the pictures version of the Pyramids and Palm Trees task as a predictor of reading performance. They reported, as we have, that the pictures version of the task was not informative as a predictor. Binney et al (2016) did, however, demonstrate that there was a relationship between scores on the written word version of Pyramids and Palm Trees and irregular word reading accuracy. Their conclusion was that irregular word reading deficits were indicative of impairment in the semantic processing of written words. From our perspective, and following the logic of both of the theoretical models outlined in the introduction, we argue that their account is unsatisfactory. Firstly, the semantic system must be activated for accurate performance in either the word or the picture versions of the test, and according to connectionist models the input modality is irrelevant as long as there are no perceptual deficits. Secondly, the word version of Pyramids and Palm Trees and any test of reading aloud requires an orthographic input. Thus, from a DRC standpoint, it is possible that performance on both tasks could be adversely affected by damage to the orthographic lexicon - that is to say that poor scores on the word version of PPT do not have to reflect semantics at all. By this logic, it is the null relationship between picture PPT and irregular word reading that is the key finding.

It is worth noting at this point that there is an alternative method by which the predictability of spelling to sound conversion can be categorised - the concept of *consistency* (Glushko, 1979). Consistency refers to the way in which a word ending is pronounced across all members of a set of words that share spelling. An inconsistent word is one for which the same word ending is converted to different phoneme outputs in different words (e.g. the -ead ending could rhyme with head or with plead). According to Plaut et al (1996), and several other papers by those who favour the Triangle account (e.g. Woollams et al, 2016), it is consistency that is of theoretical importance, not regularity. In spite of this, Woollams et al (2007) described their stimuli as being regular or irregular. This is important for our purposes because a) some of the regular words in the stimulus set could be considered to be inconsistent and b) some of the irregular words could be considered consistent. We used the English Lexicon Project database (Balota et al, 2007) to generate orthographic neighbours for

the stimuli in Woollams et al's (2007) lists and determined the pronunciation for those words which shared the word body ending. We considered words are consistent if all the neighbours rhymed with the target word, and inconsistent if the pronunciation did not match at least two of the listed neighbours. When we omit the words that fall into category a) or b) above, reading aloud accuracy increases slightly for both NJ and JD but the overall pattern of performance does not change. The regular words that our cases read incorrectly in the earliest sessions were those with more than one word body pronunciation in English; the irregular words that the patients read correctly until the later sessions were those with consistent pronunciations.

3.2.3 Progression of lexical decision performance

Rogers et al (2004) reported an interaction between target frequency and orthographic typicality such that unusual spelling was of greater detriment to low than high frequency words, and that this pattern was exacerbated when semantic deficits were more severe. An illustration of the pattern observed in a group of semantic dementia cases by Rogers et al (2004) is provided in Figure 6. The basic pattern for semantic dementia cases is claimed to be an interaction – specifically a greater number of errors with low frequency words presented alongside a non-word foil that was higher in orthographic typicality, but no such pattern with high frequency words.

The pattern for JD is presented in the upper panel of Figure 6. In session 1, JD showed precisely the pattern of performance predicted by Rogers et al (2004) for *mild* semantic impairment. Her accuracy good for high frequency targets irrespective of orthographic typicality. For low frequency items she was adversely affected when non-word foils were higher in bigram frequency than the targets [X^2 (1) = 8.154, p < .01]. In session 5, JD showed precisely the pattern of performance predicted by Rogers et al (2004) for *severe* semantic impairment. Accuracy was good when the word was higher in bigram frequency than the non-word, and identical for targets high or low in word frequency. Accuracy was impaired for targets lower in bigram frequency than the non-words in the pair whether the target word was high [X^2 (1) = 31.151, p < .001] or low [X^2 (1) = 101.184, p < .001] in frequency. Between these sessions, however, the pattern of performance was changeable, with the orthographic typicality effect disappearing altogether in the third session with JD. In summary then, the critical presence of an orthographic typicality effect for low frequency items is by no means stable across disease progression for JD.

The pattern for NJ is presented in the lower panel of figure 6 – again, as with JD, for high frequency items performance is mildly impaired and this remains stable over testing sessions. For low frequency targets performance remained only relatively mildly impaired throughout the data collection. Interestingly though, and unlike JD, NJ's performance never showed a significant effect of orthographic typicality on lexical decision accuracy, irrespective of target frequency.

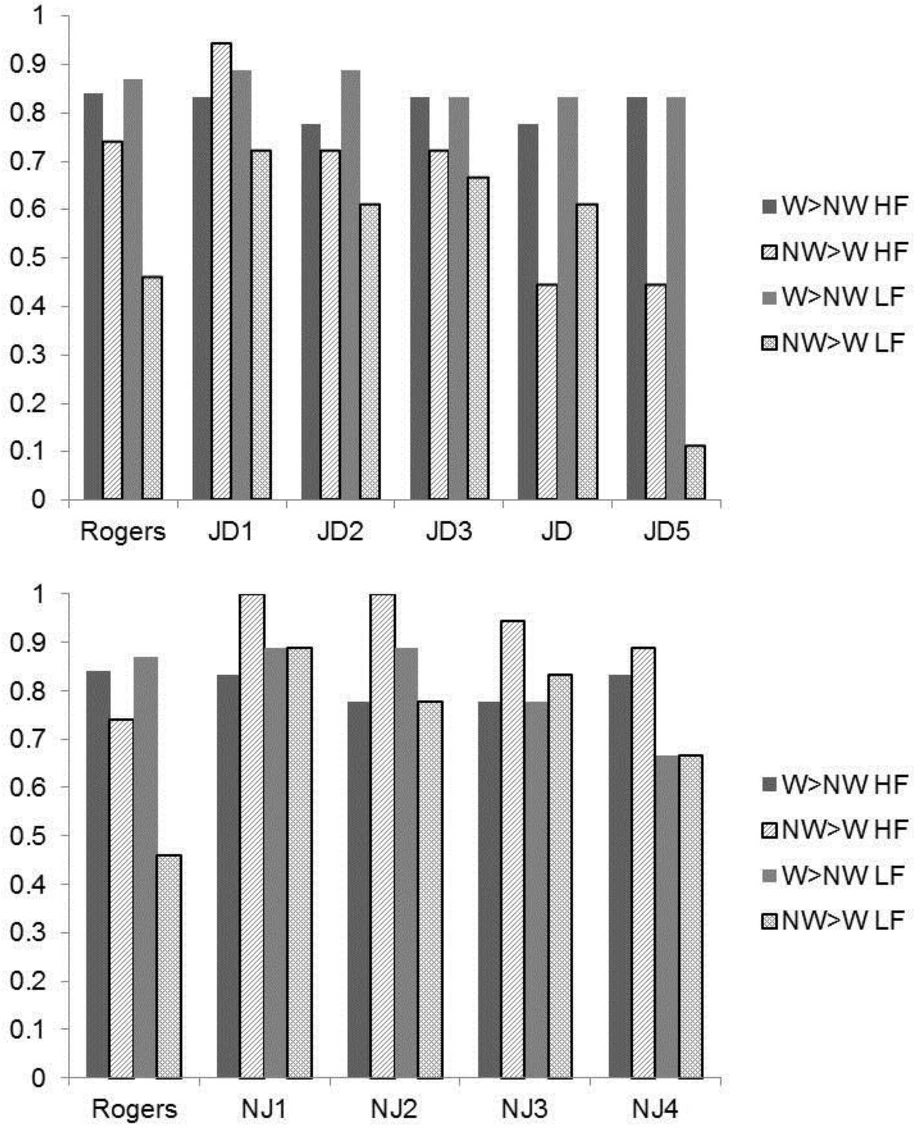


Figure 6 - Percentage accuracy for two-alternative forced choice lexical decision responses in trials where the target varied in frequency and orthographic typicality. HF = High frequency target. LF = Low frequency target. W>NW = word higher in bigram frequency than non-word. NW>W = non-word higher in bigram frequency than word.

3.2.3.1 Relationship between semantics and lexical decision performance

According to the Triangle model, **lexical decision** is contingent on the semantic system (Plaut, 1997). The DRC model does not predict a systematic relationship between semantics and lexical decision. We examined the nature of this relationship in our data by considering a) whether JD and NJ scored the same in lexical decision tasks when their semantic ability was similar, b) whether the size of any decline in semantic performance matched the size of the decline in lexical decision and c) whether semantics and lexical decision declined in parallel.

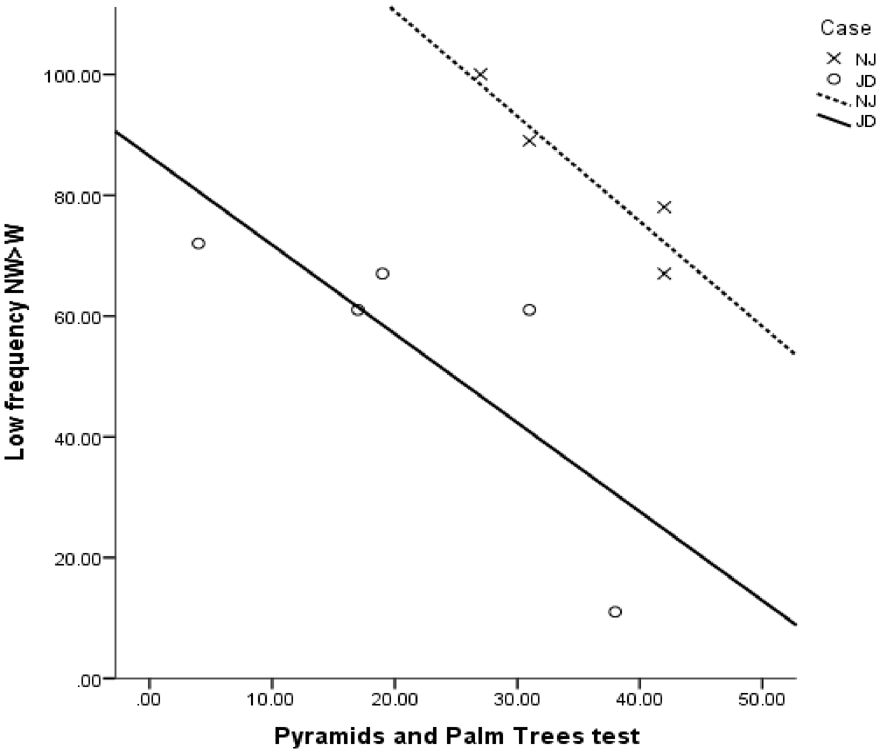


Figure 7 - Percentage correct in lexical decision trials where a low frequency word was paired with a high typicality non-word, plotted against percentage errors on the Pyramids and Palm Trees test for each case. Scores were provided in different test sessions over time, with an increasingly greater number of errors in the Pyramids and Palm Trees test as the disease progressed. Therefore data points towards the right of the x axis are from later test sessions.

Figure 7 plots the score attained on the Pyramids and Palm Trees test against accuracy in identifying low frequency words in pairs of stimuli where the non-word was higher in bigram frequency (the trials which are theoretically most reliant on semantic activation). It is important to note that even in test sessions where the semantic ability of JD and NJ did not differ, their performance on the items in the Rogers et al (2004) task most likely to require semantic activation *was* significantly different. For example, NJ was more accurate in the low frequency NW>W trials than JD (90% versus 61%) when they both scored 69% on Pyramids and Palm Trees [X^2 (1) = 21.192, p < .001]. It is also evident that there were differences between our cases in terms of the impact of their semantic decline on lexical decision performance. Finally, in pairs of test sessions where semantic performance declined significantly for each patient, success in identifying the word in the low frequency NW>W pairs did not also decline.

3.3 MRIs data results

During the time period of the extensive longitudinal behavioural testing, both patients received two structural MRI. For case JD, two scans were undertaken, corresponding to periods 3 and 5 in our behavioural testing, and for case NJ two scans were undertaken corresponding to periods 1 and 3 – in both cases we were interested to explore the profile of neurodegenerative change in each case across the two scan periods. Importantly (as elaborated above) with respect to the issue of surface dyslexic performance, it is key to point out that NJ at period 1 was *not* surface dyslexic despite having a semantic memory impairment, but *was* surface dyslexic at period 3 (when our second MRI scan was undertaken). On the other hand, JD at time period 3 (corresponding to our first MRI scan) was semantically impaired, particularly via the written word, and had poor LF irregular word reading and a mild impairment of lexical decision. By time period 5 (corresponding to our second MRI scan) JD was severely impaired at all word reading and lexical decision. The purpose of our exploratory comparative structural MRI analyses is to determine what key neural changes occurred over these two time periods and reflect upon them with respect to the behavioural impairment changes that also occurred over these periods. For the sake of brevity we focus on left hemisphere changes.

3.3.1 NJ

As you will recall, NJ's two MRI scans covered behavioural testing sessions 1 and 3 (see Figure 8) – during which he changed from having no surface dyslexia to becoming impaired at reading irregular words (and lexical decision). In the same period, NJ's performance on the picture version of the Pyramids and Palm Trees task had decreased to chance levels.

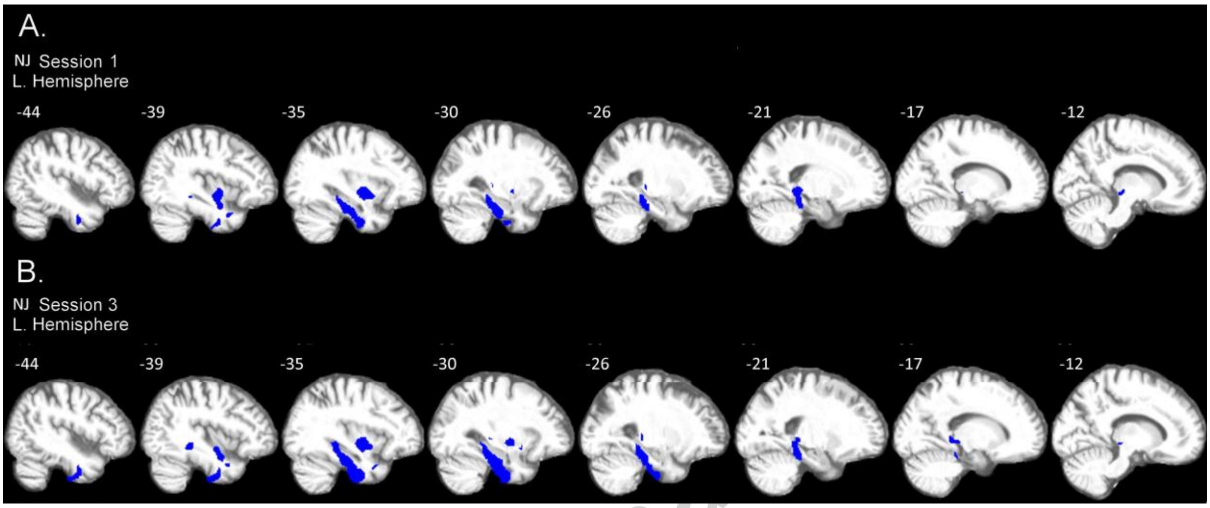


Figure 8: Left hemisphere grey matter volume in NJ (<controls, cluster level correction p(FWE) < 0.05, k=525) in sessions 1 and 3. Slice x coordinates displayed.

In Scan 1, NJ displayed a focal region of atrophy in the left anterior fusiform gyrus, which extends medially into the parahippocampal gyrus and dorsally into the insula (see figure 8A and Table 3). By Scan 2, atrophy has spread ventrally, into the inferior temporal gyrus of the ATL, as well as extending anteriorly, presenting a small cluster in the superior temporal gyrus of the ATL (see figure 8B, sagittal slice x = -39 through -26).

NJ	Cluster	K extent	Cluster peak	x	y	z

Scan 1	1.	2311	L. Fusiform gyrus	-32	-30	-23
				-33	-15	-35
			L. Parahippocampal gyrus	-21	-35	-18
			L. Insula	-37	-4	-7
Scan 2	1.	2848	L. Fusiform gyrus	-32	-30	-23
				-32	-15	-38
			L. Parahippocampal gyrus	-21	-35	-18
			L. Insula	-37	-4	-7
			L. Inferior temporal gyrus	-36	-14	-47
			L. Superior temporal gyrus	-34	9	-33

Table 3: Grey matter volume atrophy (cluster peaks) for NJ (< controls, cluster level correction p(FWE) < 0.05, k=525) across both scanning periods.

3.3.2 JD

To reiterate, JD’s two MRI scans covered behavioural testing sessions 3 and 5 (see Figure 9) – during which she changed from having poor LF irregular word reading and a mild impairment of lexical decision, to becoming severely impaired at all word reading and lexical decision. In the same period, her accuracy in the Pyramids and Palm Trees task had decreased considerably (with performance on the written version already at chance).

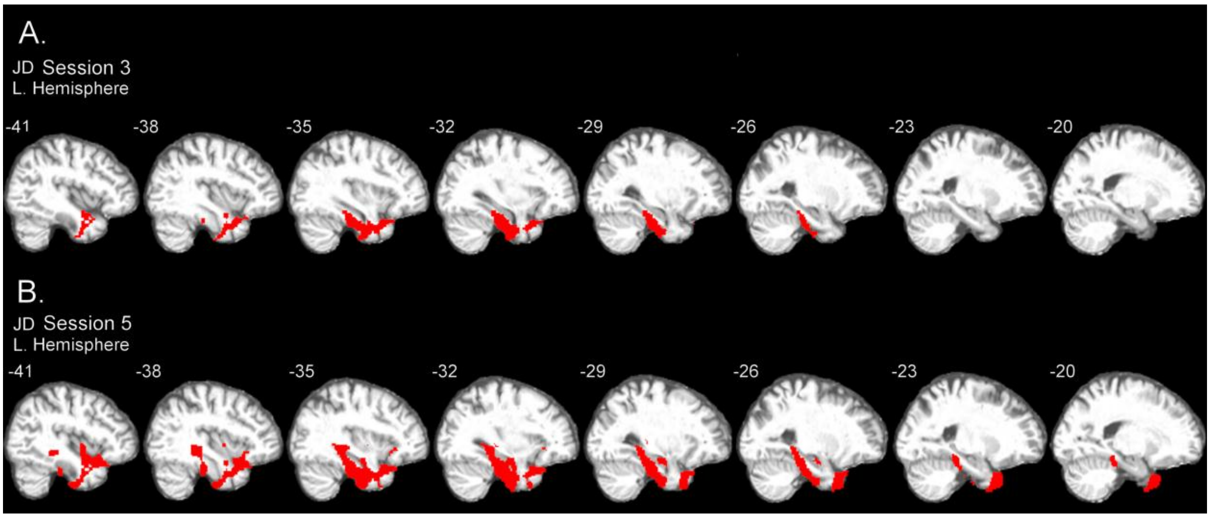


Figure 9: Left Hemisphere grey matter volume in JD (<controls, cluster level correction p(FWE) < 0.05, k=525) in sessions 3 and 5. Slice x coordinates displayed.

Exploratory analyses of MRI scan 1 for JD indicated that she displayed a focal region of atrophy in the left anterior fusiform gyrus which extends into the temporal pole/inferior temporal gyrus. It is notable that the peak of this cluster of atrophy is in the same region as the cluster displayed by NJ at time 1 and time 2; however JD's cluster extended more ventrally and anteriorly. It also does not extend into the insula cortex. By scan 2, JD was displaying reduced grey matter volume in a much more extensive region of the temporal pole (see figure 9, sagittal slice x = -26 through -20) and posteriorly in the fusiform gyrus and a smaller region in the left anterior insula (see Table 4).

JD	**Cluster**	**K extent**	**Region**	**x**	**y**	**z**
Scan 1	1.	2340	L. Fusiform gyrus	-32	-30	-23
				-33	-17	-35
			L. Inferior temporal gyrus	-35	-14	-47
			L. Superior Temporal gyrus	-37	4	-33
Scan 2	1.	4825	L. Fusiform gyrus	-32	-30	-23

				-36	-14	-47
			L. Inferior temporal gyrus	-32	-14	-39
				53	18	-6
				65	11	-6
	3.	2558	L. Superior Temporal Gyrus	-35	12	-29
				-21	15	-39
			L. Anterior insula	-45	3	-8

Table 4: Clusters of grey matter volume for JD (< controls, cluster level correction p-(FWE) < 0.05, k=525) across scanning both periods.

In summary, our exploratory neuroimaging analyses suggest that for both cases, early semantic deficit was associated with atrophy in the region of the fusiform gyrus/ parahippocampal gyrus. The findings are strikingly similar to that of the study by Bright et al (2008), who found early fusiform atrophy in two patients with a diagnosis of SD.

The emergence of surface dyslexia by Scan 2 in NJ implicates the superior temporal region of the ATL as well as increasing atrophy in more ventral regions of the temporal lobe. Such regions were already atrophic for JD in the first scan (by which she was also surface dyslexic), with greater atrophy in similar regions by Scan 2 (when JD was even more impaired at reading). On balance, our scanning explorations suggest that it may well be neurological degeneration occurring both inside and outside the frank ATL regions that are in fact key to the surface dyslexia symptoms experienced by both our semantic dementia cases. Furthermore, this is the second longitudinal study (in addition to Bright et al, 2008) to implicate that atrophy in regions outside of the temporal pole can result in deficits in semantic processing early on in disease progression.

4. General Discussion

In this paper we have presented the most comprehensive longitudinal study of reading performance in semantic dementia to date. Our main findings can be summarised as follows. First, as expected in these kinds of presentations, semantic performance declined over the course of the study. Secondly, we have confirmed that phonological performance is not longitudinally detrimentally affected in Semantic Dementia. This confirms what has previously been suggested in the literature (Hodges et al, 1992; Jefferies et al, 2005; Knott et al, 1997; 2000; Snowden et al, 1996). Importantly, this study has also shown that, a) semantic performance does not predict irregular word reading performance, b) phonological performance does not predict non-word reading performance and c) semantic performance does not predict lexical decision performance (at least in our cases). These last three findings are not compatible with the Triangle account of reading (Plaut et al, 1996) but can be easily accommodated within a dual route framework. We also assessed the progression of the atrophy in both our cases over the test period. Our findings suggest that it is not just the degree of damage to the ATL that is predictive of irregular word reading performance, as Triangle accounts would suggest; instead it appears that the emergence of damage to posterior and inferior regions of the temporal lobe are perhaps more appropriate indicators of deficits in reading aloud this type of stimulus. We will revisit this later.

Blazely et al (2005) suggested that reading performance in Semantic Dementia could potentially be characterised as three stages. Initially, SD results in impaired comprehension without detriment to reading accuracy. In stage 2, irregular word reading is poor, but regular and non-word reading is intact. In the final stage, non-word reading becomes impaired too, and in some cases regular word reading is also affected. The existence of these stages was supported by examination of the cases reported in Woollams et al (2007). Here, we have provided additional support for Blazely et al (2005) by tracking the reading performance of the same cases as their SD progressed. Both JD and NJ scored below normal on the Pyramids and Palm Trees task from the first test session, and both were impaired in irregular word reading. This is the classic surface dyslexic pattern. NJ later developed a significant impairment for non-word reading as well (moving into stage 3) while JD did not. In sum, our data suggest that individual SD cases do indeed pass through the stages proposed by Blazely et al (2005). It appears from the data reported here that NJ's progression to stage 3 had little to do with semantic or phonological ability, in contrast to the primary systems hypothesis.

Of further interest is the relationship between these reading deficits and semantic and phonological ability. In both of the cases described here, a decrease in semantic performance was coupled with a decrease in irregular word reading success. However, the absolute level of semantic impairment did not indicate the severity of the irregular word reading difficulty. For example, at a time when JD scored 69% on Pyramids and Palm Trees, she read 48% of low frequency irregular words correctly. At a time when NJ scored 69% on Pyramids and Palm Trees, he was accurate for 64% of low frequency irregular words. This would be **compatible with** the DRC framework, which posits no predictable relationship between semantics and irregular word reading (Coltheart et al, 2011). The Triangle account can also explain this dissociation by appealing to differences in the importance of semantics in pre-morbid irregular word reading (Woollams et al, 2007). Importantly the relationship between the decline in both semantic performance and irregular word reading ability for our cases does not match this pattern. It is our position that our findings are problematic for Woollams et al's (2007) hypothesis, because in individuals for whom semantics are relatively less important in irregular word reading, a larger decline in semantic performance would be required to cause the same detriment in reading performance. That was not the case in our data - in fact the patient for whom irregular word reading initially appeared less reliant on semantics was *more* affected by the degradation of the semantic system. In relation to non-word reading, the Triangle position is that phonological ability is the key factor (e.g. Harm &

Seidenberg, 2001). Again, our data are problematic for this hypothesis. Eventually one of our cases (NJ) performed significantly below the normal range in non-word reading, but showed little serious deficit on phonological tasks (i.e., digit span, non-word repetition). This pattern, too, can be explained easily by a dual route model, which proposes a non-lexical route for non-word reading. Damage to this reading specific system could adversely affect non-word accuracy without altering general phonological ability (see Nickels et al, 2008 for a simulation of non-word reading).

One puzzling finding is that JD's non-word reading accuracy was consistently higher than her accuracy for low frequency regular words. This is a problem for both the DRC account and the triangle account alike, given that in either model the process for computing the pronunciation of a non-word can be applied successfully to regular words as well. On closer inspection, this pattern was evident in the more severely impaired groups of patients described by Woollams et al (2007) but they did not offer any kind of explanation for this. It is possible that this pattern is not a product of the reading process itself, but instead reflects differences in the scoring of responses to the stimuli. For example, "sour" is considered a regular word – yet the "our" ending is pronounced differently in "pour" than it is in sour, and the "ou" vowel cluster can also be pronounced as it is in "soul" or in "through." This means that there are many ways that "sour" *could* be converted into speech output of which only one is deemed acceptable. Indeed, most of the errors made by JD (and NJ, for that matter) in regular word reading were on the items for which the spelling to sound conversion differs across different words. For non-words, on the other hand, there *is* no correct pronunciation by virtue of the fact that the non-word does not exist in the language. Hence any legitimate conversion of graphemes to phonemes will score on a non-word trial (indeed the appendix of Woollams et al, 2007, provides several pronunciations for

each of the non-words) so there are a greater number of potential scoring responses than there are to a regular word. In fact, there have been several studies assessing monosyllabic and disyllabic non-word reading in healthy participants that have demonstrated empirically that a range of pronunciations are not only possible in theory, but are produced in practice to varying degrees (Mousikou, Sadat, Lucas & Rastle, 2017; Pritchard, Coltheart, Palethorpe & Castles, 2012). The *H statistic* (Shannon, 1949) has been used to quantify this variability in non-word pronunciation. It would be perhaps be advisable, in future, to take account of this factor in the selection of non-words so that regular word and non-word reading performance is more directly comparable.

The remainder of our behavioural findings relate to the role of semantic ability in successful lexical decision. To reiterate, the Triangle account argues that the semantic system is integral for word recognition such that a semantic deficit necessarily equals poor lexical decision performance; the DRC **account** does not implicate semantics in lexical decision and hence suggests no systematic relationship between the two. Here too our data is supportive of **the DRC account**. Firstly, Rogers et al (2004) reported a frequency x orthographic typicality interaction in their two-alternative forced choice lexical decision task such that accuracy for low frequency words was lower when the spellings were unusual of the English language than when spellings were common. Rogers et al (2004) argued that these items required a greater degree of semantic input to support their recognition, and hence this interaction would be exacerbated in semantic dementia cases. In the current study JD showed the predicted interaction, but NJ did not. Secondly, the absolute level of semantic impairment was not predictive of lexical decision performance. NJ was significantly better at identifying the word in low frequency NW>W pairs than JD when their semantic scores were similar. Thirdly, significant declines in semantic ability did not co-occur with significant decreases in lexical decision accuracy. Additionally, significant decreases in lexical decision accuracy were observed following smaller decreases in semantic ability for JD than for NJ. Taken as a whole, these findings pose a problem for a Triangle account in which word recognition is underpinned by the semantic system (Plaut, 1997; **Plaut & Booth, 2006**) - at the very least the relationship between semantics and word recognition is not as straightforward as previously argued. Again, the findings relating to lexical decision performance reported in the current study can easily be accommodated in the DRC model. Coltheart et al (2010) demonstrated that the interaction between frequency and orthographic typicality reported by Rogers et al (2004) could be simulated by damaging only the orthographic lexicon. As the semantic system is not required to reach a lexical decision in DRC, the level of semantic deficit is not a predictive factor in lexical decision performance.

As a slight aside, the lexical decision task that we, and Rogers et al (2004), conducted manipulated orthographic typicality as measured by bigram frequency. However, it is worth noting that there are alternative methods for quantifying the "typicality" of a given word in the language. One such metric is the number of orthographic neighbours (N, Coltheart et al, 1977), and it is the case that the typical items in Rogers et al's (2004) stimuli have higher N than the atypical items. It has been shown that N may be an influential factor in word recognition (e.g. Andrews, 1989; 1992; Forster & Shen, 1996), particularly for older adults (Balota, et al, 2004). It is unclear how effects of N (a lexical variable) can be accommodated in models which posit that word recognition is achieved on the basis of activity in the semantic system (e.g. Plaut, 1997; Plaut & Booth, 2006). Indeed, this question has been hanging over the Triangle account for many years, at least since the work of Besner, Twilley, McCann and Seergobin (1990). We are not aware of any existing case study of reading aloud or word recognition performance in semantic dementia that has explicitly varied N to date, but

we argue that it would be worth exploring in future research. The reason is that if the pattern of N effects observed in skilled readers is mirrored in SD, even once the semantic system has been compromised, then this would be a clear indication that there are variables that have their effect at level of the orthographic lexicon (and by extension provide evidence that there *is* an orthographic lexicon). Indeed, it is possible that the interaction between familiarity and orthographic typicality reported by Rogers et al (2004) is in fact an interaction between familiarity and N - this pattern would further support our claims that the key determinant of lexical decision performance in SD is not semantic damage, but progression of that damage to the orthographic lexicon too.

The analysis of the MRI scans collected during the course of this study demonstrated somewhat different patterns of atrophy unfolding over time in our two cases. Of particular interest to the current study is the relative damage to areas relating to inferior temporal lobe (especially fusiform gyrus) in which both participants displayed atrophy cluster peaks in the first session. This atrophy coincided with semantic difficulties in both patients, paralleling the longitudinal study by Bright et al (2008). Furthermore, Binder et al (2009) and Taylor et al (2013) have posited that this area could play some involvement in semantic processing on the basis of neuroimaging data. Whilst these semantic difficulties coincided with severe deficits in reading of irregular words and in word recognition tasks in the first session for JD, this was not the case for NJ. In the framework of the triangle model, semantic ability is the key determinant of success in irregular word reading and word recognition; however, our behavioural results do not reflect this perspective. Furthermore, it is interesting to note that JD showed damage to more anterior areas of the temporal lobe at time point 1, particular in the superior and inferior temporal regions while NJ did not. JD exhibited the more severe deficits in reading of irregular words and in word recognition tasks in this first MRI session. In the middle phase of the testing period, NJ began to show damage to anterior temporal lobe areas that was roughly equivalent in location to that in JD (superior ATL) and atrophy in more posterior regions of the temporal lobe began to spread inferiorly. It was at this time that NJ's irregular word reading became far worse than JD’s. This region has been considered analogous to the orthographic lexical processing component of the DRC model (Taylor et al, 2013). While the DRC model does not make any explicit predictions about the brain regions that underpin reading processes, it *does* predict that irregular word reading can be spared in the face of considerable semantic damage, because irregular word reading does not require the semantic system. Only damage to the orthographic lexicon necessarily leads to irregular word reading deficits. That is, the semantic system and the orthographic lexicon are separate functional systems in DRC and hence they are presumably separate neural systems as well. Our MRI results suggests that lexical information can be represented independently in both inferior temporal regions and more superior portions of the ATL, and semantic representations may exist outside of frank ATL within the region of the anterior fusiform. It is pertinent to note, at this point, that our MRI results do not seem to implicate damage to the visual word form area (Cohen et al, 2000), an area of cortex that has previously been suggested as a candidate for the orthographic lexicon, in the development of surface dyslexia. In neither of our cases was the visual word form area shown to be atrophied in the scans that coincided with behavioural sessions in which a surface dyslexic reading pattern was observed. It is possible, of course, that the spread of the atrophy was such that pathways to the visual word form area were no longer available to JD or NJ prior to the development of surface dyslexia but we cannot provide evidence which speaks to this issue from the current data set, and addressing this question would be a useful avenue for future imaging work. That said, we *have* implicated damage to the inferior temporal gyrus in surface dyslexia. These are areas which have previously been argued to have the function of orthographic long term memory (e.g. Rapp, Purcell, Hillis, Capasso & Miceli, 2016; Tsapkini & Rapp, 2010) and/or have a role in word comprehension over and above general semantic access (Bonilha, Hillis, Hickock, den Ouden, Rorden & Fridriksson, 2017). **We therefore tentatively suggest that the presence of surface dyslexia is not predicated on the basis of damage to a general semantic system (the ATL, parahippocampal and anterior fusiform regions), nor is it due to damage to a visual word form area, but, rather, that surface dyslexia may occur because the interface between orthographic word forms and other stores of information about these items (be that semantic or phonological) is disrupted in some way. Importantly we argue that this disruption is specifically damaging to the reading process and not a general impairment of semantics that will equally affect all tasks requiring semantic processing. Again, the current dataset cannot rule out the possibility that there is damage occurring in the progression of SD which is preventing the passage of information along the processing stream - we merely offer a simple explanation for the findings we have described which can be specifically tested by future studies.**

We consider that our findings here contribute to a growing body of neuropsychological and imaging evidence that does not match the predictions of the triangle model. Our study also provides support to both the studies of Binder et al, (2016, in stroke cases exhibiting surface dyslexia) and Binney et al (2016, in semantic dementia cases),which have indicated that it is the level of atrophy to areas considerably more posterior to the ATL that are related to irregular word reading performance. It remains a challenge in the future to unpack the link between these posterior lesions in stroke and in semantic dementia cases and surface dyslexia - but we would tentatively suggest, from a behavioural perspective at least, that a DRC account offers a better fit for the data.

5. *References*

Andrews, S. (1989). Frequency and neighbourhood effects on lexical access: activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition, 15 (5),* 802-814.

Andrews, S. (1992). Frequency and neighbourhood effects on lexical access: lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition, 18 (2),* 234-254.

Ashburner J. (2007) Fast diffeomorphic image registration algorithm. *NeuroImage*. *38(1),* 95-113.

Ashburner J, & Friston KJ. (2000). Voxel-based morphometry—the methods. *Neuroimage, 11,* 805–821.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*, 283-316.

Balota, D.A., Yap, M., Cortese, M., Hutchison, K., Kessler, B., Loftis, B., Neely, J., Nelson, D., Simpson, G., & Treiman, R. (2007). The English lexicon project. *Behavior Research Methods, 39(3),* 445-459.

Besner, D., Twilley, L., McCann, R., & Seergobin, K. (1990). On the association between connectionism and data: Are a few words necessary? *Psychological Review, 97,* 432-446.

Binder, J.R., Pillay, S.B., Humphries, C.J., Gross, W.L., Graves, W.W. & Book, D.S. (2016). Surface errors without semantic impairment in acquired dyslexia: a voxel-based lesion-symptom mapping study. *Brain, 139(5),* 1517-1526.

Binney, R.J., Henry, M.L., Babiak, M., et al (2016). Reading words and other people: A comparison of exception word, familiar face and affect processing in the left and right temporal variants of primary progressive aphasia. *Cortex, 82,* 147-163.

Bishop, D.V.M. (1989). *Test for The Reception of Grammar* (2nd ed.). Age and Cognitive Performance Research Centre, University of Manchester.

Blazely, A. M., Coltheart, M., & Casey, B. J. (2005). Semantic impairment with and without surface dyslexia: Implications for models of reading. *Cognitive Neuropsychology, 22*, 695–717.

Bonilha, L., Hillis, A.E., Hickock, G., den Ouden, D.B., Rorden, C., & Fridriksson, J. (2017). Temporal lobe networks supporting the comprehension of spoken words. *Brain,* https://doi.org/10.1093/brain/awx169

Bright, P., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2008). Longi- tudinal studies of semantic dementia: The relationship between struc- tural and functional changes over time. *Neuropsychologia, 46,* 2177– 2188.

Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole-word and analytic translation of spelling to sound in a non-semantic reader. In K. Patterson, M. Coltheart, & J. C. Marshall (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 3–34). London: Erlbaum.

Brambati SM, Ogar J, Neuhaus J, Miller BL, & Gorno-Tempini ML (2009). Reading disorders in primary progressive aphasia: a behavioral and neuroimaging study. *Neuropsychologia, 47*, 1893–1900

Caccappolo-van Vliet, E, Miozzo, M, & Stern, Y. (2004a). Phonological dyslexia without phonological impairment? *Cognitive Neuropsychology, 21,* 820–839.

Caccappolo-van Vliet, E, Miozzo, M, & Stern, Y. (2004b). Phonological dyslexia: a test case for reading models. *Psychological Science, 15,* 583–590.

Code, C, Tree, J.J., & Dawe, K. (2009). Opportunities to say 'yes': Rare speech automatisms in a case of progressive nonfluent aphasia and apraxia. *Neurocase, 15,* 445-458.

Cohen, L., Dehaene, S., Naccache, L., Lehéricy, S., Dehaene-Lambertz, G., Hénaff, M-A., & Michel, F (2000). The visual word form area: Spatial and temporal characterization of an initial stage of reading in normal subjects and posterior split-brain patients, *Brain, 123*, 291–307.

Coltheart, M. (1996). Phonological dyslexia: Past and future issues. *Cognitive Neuropsychology, 13,* 749-762.

Coltheart, M. (2004). Are there lexicons? *Quarterly Journal of Experimental Psychology, 57A*, 1153-1171.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108,* 204-256.

Coltheart, M., Saunders, S., & Tree, J. (2010). Computational modelling of the effects of semantic dementia on visual word recognition. *Cognitive Neuropsychology, 27(2),* 101-114.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12,* 482-486.

Davies, R.R., Graham, K.S., Xuereb, J.H., Williams, G.B., & Hodges, J.R. (2004). The human perirhinal cortex and semantic memory. *The European Journal of Neuroscience, 20(9),* 2441-2446.

Derouesne, J. & Beauvois, M.F. (1985). The ‘‘phonemic’’ stage in the non-lexical reading process: evidence from a case of phonological alexia. In Patterson KE, Marshall JC, and Coltheart M (Eds), *Surface Dyslexia: Neuropsychological and Cognitive Studies of Phonological Reading.* Hillsdale, NJ: Erlbaum.

Forster, K.I., & Shen, D, (1996). No enemies in the neighbourhood: Absence of inhibitory neighbourhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22(3),* 696-713.

Friston, K.J., Holmes, A., Poline, J.B., Price, C.J., & Frith, C.D. (1996). Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage, 4(3 Pt 1),* 223-235.

Gaser, C., & Dahnke, R. (2016). CAT -A Computational Anatomy Toolbox for the Analysis of Structural MRI Data. *Human Brain Mapping*

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5*, 674-691

Harley, T., Oliver, T.M., Jessiman, L.J., & MacAndrew, S.B.G. (2013). Ageing makes us dyslexic. *Aphasiology*, 27, 490-505.

Harm, M.W. & Seidenberg, M.S. (2001). Are there orthographic impairments in phonological dyslexia? *Cognitive Neuropsychology, 18*, 71–92.

Henson, R. (2005). What can function neuroimaging tell the experimental psychologist? *Quarterly Journal of Experimental Psychology, 58(2)*, 193-233.

Henson, R., (2006). Forward inference using function neuroimaging: dissociations versus associations. *Trends in Cognitive Sciences, 10(2)*, 64-69.

Hodges JR, Graham N, Patterson K. (1995) Charting the progression in semantic dementia: implications for the organisation of semantic memory. *Memory, 3*, 463–95.

Hodges, J., Patterson, K., Oxbury, S., & Funnell, E. (1992). Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. *Brain, 115*, 1783–1806.

Howard, D., & Patterson, K. (1992). *Pyramids and palm trees: A test of semantic access from pictures and words.* Bury St. Edmunds, England: Thames Valley Test Company.

Jefferies, E., Jones, R.W., Bateman, D., & Lambon Ralph, M.A. (2005). A semantic contribution to nonword recall? Evidence for intact phonological processes in semantic dementia. *Cognitive Neuropsychology, 22(2),* 183-212

Kay, J., Lesser, R., & Coltheart, M. (1992) *PALPA: Psycholinguistic Assessments of Language Processing in Aphasia*. Hove: Lawrence Erlbaum Associates.

Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology, 14,* 1165–1216.

Knott, R., Patterson, K., & Hodges, J. R. (2000). The role of speech production in auditory-verbal short term memory: Evidence from progressive fluent aphasia. *Neuropsychologia, 38,* 125–142.

Lambon Ralph, M. A., & Howard, D. (2000). Gogi aphasia or semantic dementia? Simulating and assessing poor verbal comprehension in a case of progressive fluent aphasia. *Cognitive Neuropsychology, 17,* 437–465.

Macoir, J., Fossard, M., Saint-Pierre, M., & Auclair-Ouellet, N. (2012). Phonological or procedural dyslexia: Specific deficit of complex grapheme-to-phoneme conversion. *Journal of Neurolinguistics, 25,* 163-177.

McCandliss, B. D., Cohen, L., & Dehaene, S. (2003). The visual word form area: Expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences, 7,* 293–299.

McCarthy, R. A., & Warrington, E. K. (1986). Phonological reading: Phenomena and paradoxes. Cortex, *22,* 359 –380.

McCarthy. R.A., & Warrington, E.K. (1990a). *Cognitive neuropsychology***.** San Diego: Academic Press

Mousikou, P., Sadat, J., Lucas, R., & Rastle, K. (2017). Moving beyond the monosyllable in models of skilled reading: Mega-study of disyllabic nonword reading. *Journal of Memory and Language, 93,* 169-192.

Nestor, P. J., Fryer, T. D., & Hodges, J. (2006). Declarative memory impairments in Alzheimer’s disease and semantic dementia. *NeuroImage, 30,* 1010–1020.

Nickels, L., Biedermann, B., Coltheart, M., Saunders, S. & Tree, J. J. (2008). Computational modelling of phonological dyslexia: How does the DRC model fare? *Cognitive Neuropsychology, 25,* 165-193.

Noble, K., Glosser, G., & Grossman, M. (2000). Oral reading in dementia. *Brain and Language, 74,* 48 – 69.

Noppeney, U., Patterson, K., Tyler, L.K., Moss, H., Stamatakis, E.A., Bright, P., Mummery, C., & Price, C.J. (2007). Temporal lobe lesions and semantic impairment: a comparison of herpes simplex virus encephalitis and semantic dementia. *Brain, 130,* 1138-1147.

Patterson, K., & Lambon Ralph, M.A. (1999). Selective disorders of reading? *Current Opinion in Neurobiology, 9 (2),* 235-239.

Plaut, D. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes*, *12*, 767-808.

Plaut, D., & Booth, J. (2006). More modeling but still no stages: Reply to Borowsky and Besner. *Psychological Review, 113,* 196-200.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103,* 56–115.

Playfoot, D, Izura, C., & Tree, J. J., (2013) Are acronyms really irregular? *Neuropsychologia*, *51 (9)*, 1673-1683

Pritchard, S. C., Coltheart, M., Palethorpe, S., & Castles, A. (2012). Nonword reading: Comparing dual-route cascaded and connectionist dual process models with human data. Journal of Experimental Psychology: Human Perception and Performance, 38, 1268–1288.

Rapp, B., Purcell, J., Hillis, A.E., Capasso, R & Miceli, G., (2016). Neural bases of orthographic long-term memory and working memory in dysgraphia. *Brain, 139,* 588-604.

Rogers, T., Lambon Ralph, M., Hodges, J., & Patterson, K. (2004). Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology, 21*, 331–352.

Seidenberg, M., & McClelland, J. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96,* 523-568.

Shallice T, Warrington EK, McCarthy R. (1983). Reading without semantics. *Quarterly Journal of Experimental Psychology, 35A*, 111–138.

Shannon, C. E. (1949). The mathematical theory of communication. In C. E. Shannon & W. Weaver (Eds.), The mathematical theory of communication (pp. 29–125). Urbana: University of Illinois Press.

Snowden, J. S., Neary, D., & Mann, D. M. A. (Eds.). (1996). *Frontotemporal lobar degeneration: Frontotemporal dementia, progressive aphasia, semantic dementia.* London: Churchill Livingstone.

Taylor J, Rastle K, Davis MH. (2013). Can cognitive models explain brain activation during word and pseudoword reading? A meta-analysis of 36 neuroimaging studies. *Psychological Bulletin, 139,* 766–791.

Tree, J. J. (2008). Two types of phonological dyslexia – a contemporary review. *Cortex, 44,* 698-706.

Tree, J.J. & Kay, J. (2006) Phonological dyslexia and phonological impairment: an exception to the rule? *Neuropsychologia*, *44*, 2861–2873.

Tree, J. & Kay, J. (2015). Longitudinal assessment of short-term memory deterioration in a logopenic variant primary progressive aphasia with post-mortem confirmed Alzheimer's Disease pathology. *Journal of Neuropsychology* 9(2), 184-202.

Tree, J. Kay, J. & Perfect, T. (2005). “Deep” language disorders in nonfluent progressive Aphasia: an evaluation of the “summation” account of semantic errors across language production tasks. *Cognitive Neuropsychology* 22(6), 643-659

Tree, J. J & Playfoot, D. (2015). Declining object recognition performance in Semantic Dementia - a case for stored visual object representations. *Cognitive Neuropsychology, 32 (7-8),* 412-426.

Tsapkini K, & Rapp B. (2010). The orthography-specific functions of the left fusiform gyrus: evidence of modality and category specificity. *Cortex, 46*, 185–205.

Visser, M., Jefferies, E. & Lambon Ralph, M. A. (2010). Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of Cognitive Neuroscience*, *22*, 1083–1094.

Woollams, A., Lambon Ralph, M. A., Plaut, D., & Patterson, K. (2007). SD squared: On the association between semantic dementia and surface dyslexia. *Psychological Review, 114,* 316-339.

Woollams, A., Lambon Ralph, M.A., Madrid, G., & Patterson, K. (2016). Do you read how I read? Systematic individual differences in semantic reliance amongst normal readers. *Frontiers in Psychology, 7:1757*

Highlights:

- Reading aloud of irregular words is not predicted by semantic performance in SD
- Lexical decision accuracy is not predicted by semantic performance in SD
- Surface dyslexia is not solely related to anterior temporal lobe damage
- Inferior temporal cortex is implicated in irregular word reading deficits