Edinburgh Research Explorer

Broadly speaking

Broadly speaking: Vocabulary in semantic dementia shifts towards general, semantically diverse words

Paul HOFFMAN*[a]

Lotte METEYARD [b]

Karalyn PATTERSON [c, d]

[a] Neuroscience and Aphasia Research Unit (NARU), University of Manchester

[b] School of Psychology and Clinical Language Sciences, University of Reading

[c] Department of Clinical Neurosciences, University of Cambridge

[d] MRC Cognition & Brain Sciences Unit, Cambridge

* Correspondence to:

Dr. Paul Hoffman, Neuroscience and Aphasia Research Unit (NARU), Zochonis Building, School of Psychological Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, UK

Tel: +44 (0) 161 275 7336

Fax: +44 (0) 161 275 2873

Email: paul.hoffman@manchester.ac.uk

Running header: Vocabulary in semantic dementia

Abstract

One of the cardinal features of semantic dementia (SD) is a steady reduction in expressive vocabulary. We investigated the nature of this breakdown by assessing the psycholinguistic characteristics of words produced spontaneously by SD patients during an autobiographical memory interview. Speech was analysed with respect to frequency and imageability, and a recently-developed measure called semantic diversity. This measure quantifies the degree to which a word can be used in a broad range of different linguistic contexts. We used this measure in a formal exploration of the tendency for SD patients to replace specific terms with more vague and general words, on the assumption that more specific words are used in a more constrained set of contexts. Relative to healthy controls, patients were less likely to produce low-frequency, high-imageability words, and more likely to produce highly frequent, abstract words. These changes in the lexical-semantic landscape were related to semantic diversity: the highly frequent and abstract words most prevalent in the patients' speech were also the most semantically diverse. In fact, when the speech samples of healthy controls were artificially engineered such that low semantic diversity words (e.g., garage, spanner) were replaced with broader terms (e.g., place, thing), the characteristics of their speech production came to closely resemble that of SD patients. A similar simulation in which low frequency words were replaced was less successful in replicating the patient data. These findings indicate systematic biases in the deterioration of lexical-semantic space in SD. As conceptual knowledge degrades, speech increasingly consists of general terms that can be applied in a broad range of linguistic contexts and convey less specific information.

Keywords: conceptual knowledge; spontaneous speech; semantic memory; semantic diversity; latent semantic analysis

1. Introduction

Conceptual knowledge is central to the production of coherent, meaningful speech. This tight coupling between knowledge and speech is perhaps demonstrated most clearly in the syndrome of semantic dementia (SD), a progressive condition in which temporal lobe atrophy is associated with deterioration in verbal and non-verbal concepts (Hodges and Patterson, 2007; Bozeat et al., 2000). Although this loss of knowledge affects all semantically-driven behaviours, verbal and non-verbal, the impact on language is particularly pronounced. Word-finding difficulty is the primary presenting symptom in most patients and remains a key feature throughout the progression of the disorder (Lambon Ralph et al., 2001; Gorno-Tempini et al., 2011). Consistent with a general deterioration of conceptual knowledge, however, anomia in SD is not really a word-*finding* difficulty so much as a word-*knowing* difficulty. Unlike many patients with aphasia, SD cases show little benefit of phonological cues, suggesting that the problem is not at the lexical or phonological stages of word retrieval (Jefferies et al., 2008). Instead, naming problems in SD are influenced by precisely the same factors that affect the patients' performance in other semantic tasks. The names of objects are more likely to be retrieved if they are high in frequency and if the objects they refer to are highly familiar (Lambon Ralph et al., 1998) and if they are typical of their taxonomic category (Woollams et al., 2008). These factors also influence success in verbal and non-verbal comprehension (Jefferies et al., 2009; Adlam et al., 2006; Rogers et al., 2004b; Funnell, 1995) as well as in non-verbal expressive tasks like drawing and picture copying (Bozeat et al., 2003). Other studies have found a direct correspondence between naming ability for particular objects and knowledge for the same items in other tasks. Patients' ability to retrieve a word is predicted by their ability to give a definition of it (Lambon Ralph et al., 1999) and to respond appropriately to its referent in non-verbal tests (Garrard and Carroll, 2006). All of these findings support that view that anomia in SD is a direct consequence of deterioration in the underlying pan-modal concepts that are necessary for meaningful speech. This conceptual degradation has been linked to progressive atrophy in anterior temporal regions (Mion et al., 2010; Nestor et al., 2006). This region makes a critical contribution to the representation of verbal and non-verbal concepts (Pobric et al., 2010; Visser and Lambon Ralph, 2011; Marinkovic et al., 2003).

While the majority of studies have focused on single word production in SD, some have investigated the impact of semantic deterioration on connected speech. Although patients exhibit subtle abnormalities in the syntactic structure of their speech (Meteyard &

Patterson, 2009, this issue) they are for the most part able to produce fluent and grammatically well-formed sentences. Speech in SD is, however, characterised by dramatic reductions in content. Patients use words from a more restricted vocabulary, with a tendency towards over-reliance on highly familiar, 'light' words (Meteyard and Patterson, 2009; Ash et al., 2006; Wilson et al., 2010; Bird et al., 2000; Patterson and MacDonald, 2006). Bird et al. (2000) analysed the vocabulary used by SD patients when describing the Cookie Theft picture. Used commonly in aphasia assessment, this is a drawing of a domestic scene in which two children surreptitiously pilfer biscuits while their mother washes up. They found that, relative to the descriptions of healthy controls, patients were more likely to use words that were high in frequency and low in imageability. The shift towards higher frequency words is not surprising, since patients display better comprehension of high frequency words across a range of tasks (Jefferies et al., 2009; Bozeat et al., 2000; Funnell, 1995). The preponderance of low imageability words is more puzzling, since most SD patients show *poorer* comprehension of these words, relative to more concrete terms, in formal testing (Jefferies et al., 2009; Hoffman and Lambon Ralph, 2011). Bird et al. argued, however, that the apparent preference for more abstract words was a direct consequence of the loss of low frequency words from the patients' vocabularies. Within the set of words typically used to describe the Cookie Theft picture, frequency and imageability are correlated such that higher frequency words tend to be less imageable. Therefore, as SD patients came to rely predominantly on higher frequency words, a concomitant reduction in imageability occurred. Bird et al. modelled this effect by creating an artificial Cookie Theft description based on those of healthy controls, and progressively degrading it by removing the low frequency terms and replacing them with higher frequency words. The engineered descriptions exhibited the same shift in vocabulary towards the more abstract end of the imageability scale. As well as providing insights into the properties of connected speech in SD, this study illustrated the importance of considering the relationships between different lexical-semantic variables when assessing their effects in speech production.

Most studies of connected speech in SD have used picture description to elicit speech samples (Ash et al., 2006; Bird et al., 2000; Wilson et al., 2010; Patterson and MacDonald, 2006). There are a number of advantages to this approach. Because the topic is highly constrained, there is high degree of uniformity in the speech produced by healthy individuals. This provides a reliable baseline to which the patients' speech can be compared, allowing for the use of particular words to be investigated (e.g., "over-flowing" in the Cookie Theft description; Patterson and MacDonald, 2006). The constrained topic also tends to magnify

word retrieval difficulties because patients are forced to attempt to describe objects and events that they may not recognise or fully understand (Sajjadi et al., 2012). In contrast, in free speech SD patients are often adept at masking their language impairments by avoiding topics for which they have little knowledge. A limitation of picture description, however, is that it mainly probes concrete objects and actions. It does not provide much opportunity for discourse about emotions or inter-personal relationships, which are an important element of real conversational speech, and it mainly solicits simple, present-tense constructions such as "The water is overflowing" (or, in the patients' speech, more typically "The water is coming down"; Patterson and MacDonald, 2006). Finally, the size of the speech sample elicited from each participant is usually limited to a couple of hundred words. In the present study, we analysed a more naturalistic set of speech samples that avoided some of these limitations. We analysed the words used by seven SD patients during an autobiographical memory interview (Kopelman et al., 1990; Irish et al., 2011). The advantages of this method are as follows:

1. The content of the speech was more representative of natural conversations. In the autobiographical memory interview, patients were asked to describe significant events from their lives. In addition to the concrete, sensory aspects of these events, patients often described their relationships with the other people involved and their emotional reactions.

2. Patients had more control over the topic. They were asked to describe events from various specific periods in their lives but were able to choose which events they described and which aspects of the event they focused on. This is more similar to everyday conversations, in which patients can attempt to minimise word-finding difficulties by selecting topics for which their semantic knowledge is more intact.

3. Longer samples of speech were elicited. While a picture description speech sample typically contains one or two minutes of speech, the memory interviews lasted between 30 and 60 minutes and yielded transcripts of at least 2000 words in length.

In a previous study, Meteyard and Patterson (2009) used these speech samples to explore speech errors in SD, with a particular focus on syntax and phonology. The goal of the present study was to investigate the lexical-semantic characteristics of the vocabulary used by the patients. First, we asked whether the findings of Bird et al. (2000) with respect to frequency and imageability would be replicated in more naturalistic speech. A second goal was to analyse the vocabulary of SD patients with respect to *semantic diversity*, a lexical-

semantic variable that quantifies the degree of variability in the contextual usage of words (Hoffman et al., 2011; Hoffman et al., in press). The basic principle behind this measure is the idea that all words occur in a variety of different linguistic contexts and that the relationships between the different contexts that contain a particular word provide useful insights into how that word is processed. Some words occur in a restricted set of contexts that are closely related in meaning (e.g., *spinach* almost always occurs in the context of eating and cooking). Other words can occur in disparate situations with little relation to one another (e.g., *predicament*). Hoffman et al. (in press) devised a method that uses latent semantic analysis (LSA; Landauer and Dumais, 1997; Deerwester et al., 1990) to quantify the degree of similarity between the different contexts in which a word occurs. Using this technique, each linguistic context within a large text corpus is represented as a point in a high-dimensional semantic space, such that the proximity of two contexts indicates their similarity in meaning. Semantic diversity is a measure of the distance between the various contexts that contain a particular word. Thus, words that tend to appear in a restricted set of contexts have low diversity values (e.g., *spinach* has a value of 0.99) and those that appear in a wider range of disparate contexts have high values (e.g., *predicament* has a value of 1.93, with the maximum possible values being around 2.4 for function words like *also, which* and *from*).

Semantic diversity provides a means of formally exploring an oft-reported characteristic of speech in SD: the tendency for patients to replace specific terms with more general, vague words (e.g., *thing, place, person*). This phenomenon has been observed in studies of picture naming, in which patients make superordinate errors (e.g., squirrel → "animal"; Woollams et al., 2008), as well as connected speech elicited from picture descriptions (Ash et al., 2006; Bird et al., 2000; Wilson et al., 2010), and is thought to reflect the gradual loss of specific features from conceptual knowledge. In this study, we used semantic diversity as a way of quantifying the specificity of different words. Highly specific words can only be used in a restricted set of contexts (hence have low semantic diversity) because they have highly specific referents. In contrast, more vague, general terms (sometimes referred to as "light" words) can refer to a variety of different items and can be used appropriately in a more diverse set of contexts, resulting in higher semantic diversity values. In line with the reported shift from specific to general terms, we predicted that SD patients would show a relative paucity of low semantic diversity words in their speech. This issue is also important because semantic diversity is correlated with both frequency and imageability (Hoffman et al., in press). High frequency words and low imageability words tend to be high in semantic diversity. Thus, a shift towards more use of high semantic

diversity words would result in patients producing fewer low frequency and high imageability words, consistent with the previous findings of Bird et al. (2000). In fact, as detailed later, we were able to accurately simulate the speech patterns of SD patients by removing low semantic diversity words from the speech of healthy controls. A similar simulation formed by removing low frequency words was less successful, suggesting that semantic diversity, rather than frequency, may be the dominant factor in determining which words patients with SD use in spontaneous speech.

2. Method

2.1 Participants

The speech samples reported here were elicited from seven patients with a clinical diagnosis of SD who took part in a study on autobiographical memory (Irish et al., 2011). Patients were identified through the Memory and Cognitive Disorders Clinic or the Early Dementia Clinic at Addenbrooke's Hospital, Cambridge, UK. They were diagnosed based on cognitive and neuroradiological criteria described by Hodges et al. (Hodges et al., 1992). Patients had a mean age of 64 years (s.d. = 6.8) and had spent an average of 13.6 years in formal education (s.d. = 3.2). Scores on standard neuropsychological tests are provided in Table 1, along with published norms from healthy individuals. All of the patients showed the typical pattern of cognitive deterioration in SD: impaired performance of verbal and non-verbal tests of semantic memory, combined with relatively preserved performance in other cognitive domains (represented here by backward digit span and copying of the Rey complex figure). The patients were compared with eight control participants, matched in age and educational background to the patients, who took part in the same autobiographical memory study. They had a mean age of 60 (s.d. = 4.9) and an average of 15.6 years in education (s.d. = 3.2). All participants gave informed consent and the study was approved by the local ethics committee.

2.2 Speech sample elicitation and transcription

The method for eliciting and transcribing speech was described in detail by Meteyard and Patterson (2009) and is summarised here. Participants were interviewed using the Autobiographical Memory Interview (Kopelman et al., 1990). They were asked to describe

specific events from different periods in their life that were particularly striking or memorable. Events from four life periods were probed. Typical events included weddings, birthdays, holidays and career-related events (e.g., new job). When participants had finished giving their spontaneous description of an event, the interviewer asked questions to prompt the retrieval of more specific information (e.g., sensory experiences or emotional reactions associated with the event). Interviews were usually completed in a single session; however, two patients (BC and PS) had follow-up interviews that were also transcribed and included in the analysis. Interviews were transcribed from the original audio recordings, including any conversation that took place before or after the formal interview. Any words that could not been identified clearly from the recording were marked and excluded from all analyses.

2.3 Construction of speech corpus

The first stage in the analysis was to construct a corpus of words used by the participants and to obtain frequency, imageability and semantic diversity values for these words.

1. *Frequency:* We obtained lemma frequencies from the CELEX database (Baayen et al., 1993). These were log-transformed.
2. *Imageability:* Imageability values are less widely available than frequency values because they rely on subjective ratings. To maximise the number of words that could be included in the analysis, we obtained imageability ratings from six published databases (Coltheart, 1981; Stadthagen-Gonzalez and Davis, 2006; Bird et al., 2001; Clark and Paivio, 2004; Cortese and Fugett, 2004; Schock et al., in press). If a word was present in more than one database, an average was taken. All databases contained ratings varying between from 100 to 700.
3. *Semantic diversity:* Semantic diversity was calculated using the method described by Hoffman et al. (in press). Higher semantic diversity values indicate words that appear in a broader range of linguistic contexts.

Words were only included in the analysis if all three values were available. For plural nouns, the word was included if all three values were available for the singular form of the word. As we were principally interested in the patients' use of semantically rich, open-class words, we excluded all words that are typically used as conjunctions, prepositions, pronouns and adverbs (according to word class frequency counts in the CELEX database; Baayen et al.,

1993). Number words were also excluded. We treated morphological variants (e.g., singular, plural, verb tenses) as separate words.

3. Results

3.1 Basic characteristics of the speech sample

Table 2 shows the number of words produced by each participant. Both token and type counts are provided, although subsequent analyses were performed on token counts only. There was no significant difference in these values between patients and controls, though there was considerable variation across participants. The table also shows the number of words for which all three psycholinguistic values were available (words with data) and how many remained following the exclusion of function words (analysed words). Neither of these counts differed between patients and controls. We succeeded in obtaining all three psycholinguistic values for 81% of the tokens produced. To ensure that the rates of missing data did not differ between groups, we calculated the percentage of tokens with all three values available for each individual participant. The percentages did not differ significantly between patients and controls (SD = 80%; Controls = 81%; $t(13) = 1.40$, $p = 0.19$). Finally, Table 2 also shows the mean log frequency, imageability and semantic diversity of the analysed tokens for each participant. There were significant differences between groups for all these measures: mean frequency and semantic diversity were higher in the SD patients, while mean imageability was lower. These effects are investigated in more detail below.

3.2 Word production by frequency

We first asked how often the patients produced words of different lexical frequencies and whether this distribution differed from that seen in healthy controls. To do this, we divided our corpus of words into six frequency bands, based on their log-transformed frequency counts (obtained from CELEX). These bands are shown on the x-axis of Figure 1a. The y-axis indicates production frequency for each group in terms of the mean log production rate, which was computed in the following manner. First, we counted the number of times each participant used words belonging to each frequency band. These counts varied considerably across participants, because the length of the speech samples was highly variable (as shown in Table 2). In order to compare values across participants, we converted

each participant's count for each frequency band into a production rate by computing it as a proportion of their total number of analysed words and multiplying by 1000. This scaled the word counts such that each participant's total was 1000, allowing for direct comparison across participants. The production rates were then log-transformed. The log transformation was necessary to reduce the influence of very high frequency words that were produced many times by all participants. Throughout the paper, log-transformed production rates were used as the dependent measure. However, parallel analyses performed on untransformed production rates gave very similar results.

Log production rates for each frequency band are shown in Figure 1a. These data were analysed with a 6 × 2 mixed ANOVA that included frequency band as a within-subjects factor and participant group as a between-subjects factor. This confirmed that there was a main effect of frequency ($F(5,65) = 247$, $p < 0.001$) as well as an effect of group ($F(1,13) = 17.8$, $p = 0.001$) and a highly significant interaction ($F(5,65) = 9.93$, $p < 0.001$). Post-hoc t-tests indicated that SD patients exhibited lower production rates than controls in the three lowest frequency bands ($t(13) > 2.25$, $p < 0.05$) with the disparity being largest for the lowest frequency band. In contrast, they were more likely to produce words in the highest band ($t(13) = 3.70$, $p < 0.003$). We also conducted an additional, non-parametric chi-square analysis that compared the distribution of the total counts in the two groups. This confirmed that the distribution of word counts across frequency bands were altered in the patient group, relative to healthy controls ($\chi^2 = 184$, $p < 0.001$).

These results replicate previous findings showing that low frequency words tend to drop out of SD patients' vocabulary and that they rely increasingly on higher frequency words to communicate (Bird et al., 2000; Patterson and MacDonald, 2006; Wilson et al., 2010), consistent with the particular susceptibility of less familiar words to semantic degradation (Lambon Ralph et al., 1998; Jefferies et al., 2009).

3.3 Word production by imageability

Figure 1b shows a similar analysis of word usage for which words were divided into five bands based on their imageability. There was a fairly even distribution across bands (note that the y-axis runs from 2.0 to 2.5) indicating that, in contrast to previous picture description studies, participants were using a variety of more abstract terms as well as highly imageable words to describe their past experiences. The highest production rates were for the 300-400 band, which include a number of high-frequency verbs (e.g., *think, look, give*) and

10

nouns (e.g., *thing, place*). A 5 × 2 ANOVA performed on these data revealed main effects of imageability ($F(4,52) = 28.9$, $p < 0.001$) and group ($F(1,13) = 8.08$, $p = 0.014$) and an interaction ($F(4,52) = 2.71$, $p = 0.04$). Patients were significantly more likely than controls to produce words in the 300-400 band ($t(13) = 2.16$, $p = 0.05$) and significantly less likely to produce words in the 500-600 band ($t(13) = 3.08$, $p = 0.009$). A non-parametric chi-square analysis on the distribution of word counts in each group supported the presence of a difference between the two groups ($\chi^2 = 43.4$, $p < 0.001$).

These data appear to show a shift in the words used by SD patients, away from highly imageable words and towards more abstract terms. However, this analysis does not take into account potential differences in word frequency between the imageability bands. Bird et al. (2000) pointed out that, in their dataset, less imageable words tended to be higher in frequency than more concrete terms, such that reliance on high frequency words would cause a concomitant shift towards production of less imageable words. To explore this possibility, in the next section we considered the effects of frequency and imageability on word production simultaneously.

3.4 Word production in a frequency-imageability space

Our remaining analyses consider word production rates in the context of a two-dimensional lexical-semantic space, in which the x-axis represents word frequency and the y-axis represents imageability (see Figure 2). Following Bird et al. (2000), we generated contour plots that mapped production rates for words belonging to each part of this two-dimensional space. Maps were generated as follows. First, the space was segmented using the frequency and imageability bands reported in the previous section. Then, for each participant, we counted the number of times they produced words from each point in the space. For example, to compute the value for the bottom left point in the space, we counted the number of times a participant produced any word with a log frequency value of <1 and imageability value of <300. As before, these counts were converted to a production rate by dividing by the participant's total count and multiplying by 1000, and they were then log-transformed. The data were averaged across participants to form the group maps shown in the first upper half of Figure 2 (the simulations are described later).

At first glance, the distribution across the space is rather similar in the two groups. For both groups, the area of the space used most often is the top left-hand corner, which is home to words that are high in frequency and low in imageability. This area contains many of

the highly familiar verbs used often in normal conversation (e.g., *know, thought, ask*). There is an additional, smaller peak for higher imageability words in the 2-2.5 frequency band. This region of the space includes many common everyday nouns, such as *wife, doctor* and *office*. For both groups, there is a pronounced but unsurprising drop in word production rates towards the low frequency end of the space, particularly for less imageable words. These words were used least often by both groups.

Though the two group maps appear broadly similar, there are some apparent differences. The high-frequency, low-imageability peak appears to have expanded in the patient group, relative to controls, suggesting that the patients use these words at a higher rate than controls. In contrast, patients appear to show lower production rates in the low frequency end of the space. To investigate these potential differences further, we computed a difference plot by subtracting the control values from the patient values at each point on the grid (see Figure 3). This plot shows areas of the lexical-semantic space that patients are less likely to use than controls in blue tones and areas that the patients use more frequently than controls in orange. To assess the statistical significance of these differences, we present a plot of p-values on the right-hand side of the figure. This map was produced by performing a t-test at each point in the frequency-imageability space, comparing production rates in the patients with those of the controls.[1] This analysis indicates that patients were significantly less likely than controls to use low-frequency, high-imageability words (the bottom right corner of the space). This area of the space contains a large number of less common words, predominately concrete nouns (e.g., *sausage, beard* and *castle*). In contrast, patients were significantly more likely than controls to use the lower imageability words (300-400) from the highest frequency band. There are relatively few words occupying this point in the space so it is possible to list them in full: *find, give, go, good, look, make, new, say, see, think* and *thing*. These words were of course used frequently by controls, but significantly more frequently by the patients. This analysis replicates the findings of the picture-description task in Bird et al. (2000), here in a more naturalistic speech production task.

3.5 Relation to semantic diversity

We next turned our attention to the new variable of semantic diversity to ask whether this could help to explain the patients' altered patterns of word usage. We began by

---

[1] There is a potential problem with multiple comparisons using this technique, since a total of 30 t-tests were performed across the space. We controlled for this using the false-discovery rate approach ((Benjamini and Hochberg, 1995) All significant p-values shown in Figure 3 remained significant when the false discovery rate was controlled with $q = 0.1$.

calculating the mean semantic diversity of words occupying each point in the frequency-imageability space. These values are mapped in Figure 4, which illustrates the relationship between semantic diversity and both frequency and imageability. The low-frequency, high-imageability words tend to be the least diverse and the highest frequency, less imageable words tend to have the greatest diversity. We predicted that SD patients would be less likely to use low semantic diversity words because they convey highly specific semantic information that is likely to be degraded. In contrast, we expected them to over-use high diversity words because these convey more general information and can be used in a variety of different situations. Figure 4 suggests that this was the case: the low-frequency, high-imageability words that the patients under-used had the lowest semantic diversity values, and the high-frequency, low-imageability words that they over-used were amongst the highest in diversity. As a further test, we directly compared the difference plot in Figure 3 with the semantic diversity values in Figure 4 by computing the correlation between the two across all points in the lexical-semantic space. There was a highly significant correlation between the two sets of values ($r(29) = 0.73$, $p < 0.001$; illustrated in Figure 5). In other words, patients were more likely than controls to use words from parts of the space where the words are highly diverse and less likely to use words from areas associated with lower diversity.

3.6 Simulating the loss of low semantic diversity words and low frequency words

So far, we have seen that SD patients produce fewer low frequency words in connected speech and also that semantic diversity is a strong predictor of abnormalities in their speech patterns. Since frequency and semantic diversity are highly correlated, it is difficult to ascertain which is more influential in determining which words patients are likely to use. In this section, we directly compared two hypotheses about speech in SD: (1) that the abnormal distribution of words across lexical-semantic space exhibited by the patients was due to the absence of low semantic diversity words from their speech and (2) that the abnormal distribution was due to the absence of low frequency words. Following the approach of Bird et al. (2000), we simulated the effects of losing low diversity words and low frequency words by artificially modifying the speech samples of healthy controls. We then compared these modified speech samples to the real patient data, to determine which simulation most closely resembled the patient data, i.e., whether the speech of SD patients was best explained by loss of low diversity words or by loss of low frequency words.

13

Method: We began by dividing the speech corpus for each participant into two separate sets. We randomly assigned each token produced to one of the two sets, resulting in two equally-sized sets for each participant. The first set was used to determine the optimum parameters for the simulation. To simulate the loss of low semantic diversity words from their vocabulary, we removed all words from the controls' data that had semantic diversity values below a certain threshold. We replaced these words with the following light nouns, all of which are high in semantic diversity and frequency: *person, place(s), something, stuff, thing(s), type(s)*. We used nouns because they are by far the dominant word class among low diversity words. We repeated this procedure for a number of different thresholds. Our goal was to find the threshold at which the simulated data most closely resembled the actual patient data (specifically, we sought the threshold that minimised the mean squared difference between the simulated data and the patient data, calculated over all points of the lexical-semantic space). The optimum threshold was a semantic diversity value of 1.45 (mean squared difference = 0.395). We then wished to test how closely this simulation, formed by removing and replacing all words with a semantic diversity of less than 1.45, could replicate the real data from our SD patients. If we tested the simulation using the same data used to determine its optimum parameters, we risked over-estimating its success. Instead, we applied the pre-determined threshold of 1.45 to the second dataset and performed the same process of replacing low semantic diversity words with light nouns. We assessed how well this second, independent simulation resembled the patient data.

A similar process was conducted to simulate the effects of loss of low frequency words. All low frequency words below a set threshold were removed from the controls' data and replaced with light nouns. Again, a number of different thresholds were tested using the first dataset and the threshold that gave the closest approximation to the patient data was selected. The optimum threshold for removing low frequency words was a log frequency value of 0.85 (mean squared difference = 0.537). We applied this threshold to the second dataset in order to evaluate how closely it resembled the real data from SD patients.

Results: We computed word production rates across the frequency-imageability space for the high-diversity and high-frequency simulated data. The contour maps are shown in the lower half of Figure 2. The removal of low diversity and frequency words had relatively subtle effects on the overall distribution. Note, however, that in both simulations the rate of low frequency words is lower than in the original control data (particularly in the <1 band) and now more closely resembles the rate in the patient data. In the high-diversity simulation, there

is also a pronounced expansion in the peak for high-frequency, low-imageability words (i.e., in the top left-hand corner of the space), so that this now resembles the pattern in the patient group. This is because most of the light nouns used to replace the removed words lie in this part of the space.

To visualise the similarities and differences between the simulated data and the actual patient data, we computed difference plots by subtracting each set of simulated data from the patient data (shown in Figure 6). Both simulations were fairly successful in approximating the patient data. However, the high-frequency simulation was less successful than the high-diversity simulation. The high-frequency simulation appeared to underestimate the rate at which patients produced words in the <1 frequency band (the orange area at the bottom of the plot) and overestimated the rate of high imageability words in the 1-1.5 frequency band (the dark blue area). To identity areas that exhibited a statistically significant prediction error, we performed t-tests that compared the data from each simulation with the real patient data. As shown in Figure 6, there was only one area of significant difference for the high-diversity simulation: it predicted a somewhat lower production rate for the highest frequency words in the 400-500 imageability band than was actually observed in the patients ($p = 0.04$). Otherwise, the simulation appeared to accurately predict production rates in the patients. The high-frequency simulation made a similar error at the same point in the lexical-semantic space. More importantly, the high-frequency simulation also differed significantly from the patient data in the 1-1.5 frequency band, for the higher imageability words. The simulation predicted a significantly higher production rate for these words than was actually observed. In other words, although this model was able to successfully simulate the reduction in the very lowest frequency words in patients with SD, it over-estimated the rate at which patients used words from the second frequency band.[2] In contrast, the high-diversity model was better able to capture the reduction in production rates across both of the lower frequency bands.

4. Discussion

We investigated the properties of the vocabulary used by SD patients during a relatively unconstrained autobiographical interview. The distribution of words used by patients differed from that of healthy controls with respect to frequency and imageability.

---

[2] This was because the optimum threshold for removing low frequency words was 0.85 and no words from the 1-1.5 band were removed at this threshold. This prediction error could be combated by adopting a higher threshold; however, at higher thresholds the *overall* prediction error was even greater because these simulations drastically under-estimated production rates in the <1 band.

Patients produced fewer low frequency, high imageability words than controls, but more high frequency, low imageability words. We then explored the possibility that these changes were due to the replacement of highly specific words, which tend to be low in frequency and high in imageability, with more vague, general terms, which have the opposite characteristics. This dimension of specificity was quantified using semantic diversity, a measure of the degree to which words can be used in a variety of different contexts. Differences between the speech of SD patients and controls were strongly correlated with semantic diversity. Furthermore, we were able to simulate the patterns of word usage in SD by removing low semantic diversity words from the speech of healthy controls and replacing them with more general, high diversity terms. A similar simulation was performed in which low frequency words were replaced but this was less successful, suggesting that semantic diversity is the better predictor of word usage in the speech of SD patients.

These findings are consistent with, indeed appropriate to, the behavioural and neuroanatomical features associated with SD. The language deficit in SD is thought to be due to the gradual deterioration of core conceptual knowledge in the condition (Rogers et al., 2004a; Patterson et al., 2007), which is in turn a consequence of atrophy in anterior temporal regions that are implicated in representing such information (Pobric et al., 2010; Visser and Lambon Ralph, 2011; Marinkovic et al., 2003; Mion et al., 2010). A number of studies have shown that concepts in SD deteriorate in a graded fashion, whereby general, superordinate information about a particular concept can remain even after more fine-grained, item-specific knowledge has been lost (Adlam et al., 2006; Crutch and Warrington, 2006; Warrington, 1975; Rogers and Patterson, 2007). In addition, it is well-established that highly familiar concepts and the high frequency words that refer to them are more robust in the face of semantic degradation than less familiar concepts (Jefferies et al., 2009; Bozeat et al., 2000; Funnell, 1995). As we have demonstrated here, these factors also strongly influence which words are available to patients in conversation. There are, however, some differences between the effects formed in formal comprehension testing and those observed in the present study. Hoffman et al. (2011) investigated effects of frequency, imageability and semantic diversity in 13 SD patients using a forced-choice verbal comprehension test (e.g., is *value* most similar to *purpose*, *effect* or *price*?). Multiple regression analysis indicated that frequency and imageability were both strong predictors of which words were comprehended but that semantic diversity had no independent effect. In contrast, the present analysis suggested that semantic diversity was a *better* predictor than frequency when considering which words are produced in connected speech. This discrepancy may provide an insight into

the differing factors at play during single-word comprehension vs. connected speech production. The verbal comprehension test is a relatively direct way of assessing the integrity of particular concepts. The primary factor that determines whether a word is comprehended is the degree to which the underlying concept is represented in the patient's degraded semantic network, which in turn depends on its frequency. Computational models of the semantic system indicate that concepts encountered frequently come to dominate the structure of the semantic network, providing them with a greater robustness to damage (Simulation 5.4; Rogers and McClelland, 2004). Therefore, the strong effect of frequency in comprehension testing reflects the fact that highly frequent words are more likely to have intact semantic representations.

The situation is more complex for production of connected speech. The integrity of particular concepts still plays an important role in determining which words a patient is able to use in speech, but one must also consider which words are used to replace those that are unavailable. When a patient is unable to retrieve a particular word they will often replace it with a substitute, as this maintains the flow of speech and avoids a protracted pause. High semantic diversity words are useful as substitutes because they can be used appropriately in many different contexts and their meanings are often somewhat flexible, allowing them to stand in for a variety of other, more specific terms. This may explain why prominent semantic diversity effects are observed when connected speech production is analysed, in addition to frequency effects. As an illustrative example, consider a highly specific, low frequency noun like *cathedral*. Where unimpaired speakers would use this word, patients with SD might instead use *church, building* or *place*, depending on the severity of their semantic impairment. The psycholinguistic properties of these words are shown in Table 3. *Cathedral* is much lower in frequency than the other terms and its semantic representation is therefore more vulnerable to degradation. It also has a low semantic diversity value, indicating that it occurs in a restricted set of linguistic contexts. The remaining words, which could be used as substitutes, are all higher in frequency and in semantic diversity. *Place* is the most semantically diverse, as it can be used in a variety of situations with a variety of different referents. *Church* and *building* fall somewhere in between. In this case, the semantic diversity of the various terms also mirrors the degree of perceptual variation associated with their referents. *Cathedral* is used to refer to a specific class of buildings with a fairly rigid set of characteristics (e.g., large, ornate, made of stone). There is more perceptual variation amongst the things that people refer to as churches, still more amongst buildings and most amongst "places". It is interesting that in this case, the semantic diversity of the words is

correlated with the degree of perceptual variation present in the items they refer to. Therefore, even though semantic diversity is a measure of variation in linguistic use, it may also be informative about the degree of variation among the physical objects words are used to refer to.

Returning to Table 3, note that *place* is much less imageable than the other terms, presumably because it is hard to associate it with any particular image. This example illustrates a key finding from the study: that a shift from more specific/low frequency to more general/high frequency terms results in a tendency to use less imageable words. Therefore, the tendency for SD patients to use many high frequency, low imageability words, which is often apparent in a clinical setting, can be traced directly to the underlying deterioration of semantic knowledge that is the hallmark of this condition.

References

Adlam ALR, Patterson K, Rogers TT, Nestor PJ, Salmond CH, Acosta-Cabronero J, and Hodges JR. Semantic dementia and fluent primary progressive aphasia: two sides of the same coin? *Brain,* 129: 3066-3080, 2006.

Ash S, Moore P, Antani S, McCawley G, Work M, and Grossman M. Trying to tell a tale: Discourse impairments in progressive aphasia and frontotemporal dementia. *Neurology,* 66: 1405-1413, 2006.

Baayen RH, Piepenbrock R, and van Rijn H. *The CELEX lexical database (CD-ROM).* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania, 1993.

Benjamini Y and Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological,* 57(1): 289-300, 1995.

Bird H, Franklin S, and Howard D. Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods Instruments & Computers,* 33(1): 73-79, 2001.

Bird H, Lambon Ralph MA, Patterson K, and Hodges JR. The rise and fall of frequency and imageability: Noun and verb production in semantic dementia. *Brain and Language,* 73(1): 17-49, 2000.

Bozeat S, Lambon Ralph MA, Graham KS, Patterson K, Wilkin H, Rowland J, Rogers TT, and Hodges JR. A duck with four legs: Investigating the structure of conceptual knowledge using picture drawing in semantic dementia. *Cognitive Neuropsychology,* 20(1): 27-47, 2003.

Bozeat S, Lambon Ralph MA, Patterson K, Garrard P, and Hodges JR. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia,* 38(9): 1207-1215, 2000.

Clark JM and Paivio A. Extensions of the Paivio, Yuille, and Madigan (1968) norms. *Behavior Research Methods Instruments & Computers,* 36(3): 371-383, 2004.

Coltheart M. The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology,* 33(Nov): 497-505, 1981.

Cortese MJ and Fugett A. Imageability ratings for 3,000 monosyllabic words. *Behavior Research Methods Instruments & Computers,* 36(3): 384-387, 2004.

Crutch SJ and Warrington EK. Partial knowledge of abstract words in patients with cortical degenerative conditions. *Neuropsychology,* 20(4): 482-489, 2006.

Deerwester SC, Dumais ST, Landauer TK, Furnas GW, and Harshman RA. Indexing by latent semantic analysis. *Journal of the Americal Society of Information Science,* 41(6): 391-407, 1990.

Folstein MF, Folstein SE, and McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research,* 12: 189-198, 1975.

Funnell E. Objects and properties: A study of the breakdown of semantic memory. *Memory,* 3(3-4): 497-518, 1995.

Garrard P and Carroll E. Lost in semantic space: a multi-modal, non-verbal assessment of feature knowledge in semantic dementia. *Brain,* 129: 1152-1163, 2006.

Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF, Ogar JM, Rohrer JD, Black S, Boeve BF, Manes F, Dronkers NF, Vandenberghe R, Rascovsky K, Patterson K, Miller BL, Knopman DS, Hodges JR, Mesulam MM, and Grossman M. Classification of primary progressive aphasia and its variants. *Neurology,* 76(11): 1006-1014, 2011.

Hodges JR and Patterson K. Semantic dementia: a unique clinicopathological syndrome. *Lancet Neurology,* 6(11): 1004-1014, 2007.

Hodges JR, Patterson K, Oxbury S, and Funnell E. Semantic dementia: Progressive fluent aphasia with temporal lobe atrophy. *Brain,* 115: 1783-1806, 1992.

Hoffman P and Lambon Ralph MA. Reverse concreteness effects are not a typical feature of semantic dementia: Evidence for the hub-and-spoke model of conceptual representation. *Cerebral Cortex,* 21: 2103-2112, 2011.

Hoffman P, Lambon Ralph MA, and Rogers TT. Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, in press.

Hoffman P, Rogers TT, and Lambon Ralph MA. Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience,* 23(9): 2432-2446, 2011.

Irish M, Hornberger M, Lah S, Miller L, Pengas G, Nestor PJ, Hodges JR, and Piguet O. Profiles of recent autobiographical memory retrieval in semantic dementia, behavioural-variant frontotemporal dementia, and Alzheimer's disease. *Neuropsychologia,* 49(9): 2694-2702, 2011.

Jefferies E, Patterson K, Jones RW, and Lambon Ralph MA. Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology,* 23: 492-499, 2009.

Jefferies E, Patterson K, and Lambon Ralph MA. Deficits of knowledge versus executive control in semantic cognition: Insights from cued naming. *Neuropsychologia,* 46: 649-658, 2008.

Kopelman M, Wilson BA, and Baddeley AD. *The autobiographical memory interview*. Bury St Edmunds: Thames Valley Test Company, 1990.

Lambon Ralph MA, Graham KS, Ellis AW, and Hodges JR. Naming in semantic dementia - what matters? *Neuropsychologia,* 36(8): 775-784, 1998.

Lambon Ralph MA, Graham KS, Patterson K, and Hodges JR. Is a picture worth a thousand words? Evidence from concept definitions by patients with semantic dementia. *Brain and Language,* 70(3): 309-335, 1999.

Lambon Ralph MA, McClelland JL, Patterson K, Galton CJ, and Hodges JR. No right to speak? The relationship between object naming and semantic impairment: Neuropsychological abstract evidence and a computational model. *Journal of Cognitive Neuroscience,* 13(3): 341-356, 2001.

Landauer TK and Dumais ST. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review,* 104: 211-240, 1997.

Marinkovic K, Dhond RP, Dale AM, Glessner M, Carr V, and Halgren E. Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron,* 38(3): 487-497, 2003.

Mathuranath PS, Nestor PJ, Berrios GE, Rakowicz W, and Hodges JR. A brief cognitive test battery to differentiate Alzheimer's disease and frontotemporal dementia. *Neurology,* 55(11): 1613-1620, 2000.

Meteyard L and Patterson K. The relation between content and structure in language production: An analysis of speech errors in semantic dementia. *Brain and Language,* 110(3): 121-134, 2009.

Mion M, Patterson K, Acosta-Cabronero J, Pengas G, Izquierdo-Garcia D, Hong YT, Fryer TD, Williams GB, and Hodges JR. What the left and right fusiform gyri tell us about semantic memory. *Brain,* 133: 3256-3268, 2010.

Nestor PJ, Fryer TD, and Hodges JR. Declarative memory impairments in Alzheimer's disease and semantic dementia. *Neuroimage,* 30(3): 1010-1020, 2006.

Osterrieth P. Le test de copie d'une figure complexe. *Archives de Psychologie,* 30: 205-220, 1944.

Patterson K and MacDonald MC. Sweet nothings: Narrative speech in semantic dementia. In Andrews S (Ed.) *From inkmarks to ideas: Current issues in lexical processing*. Hove, UK: Psychology Press, 2006: 299-317.

Patterson K, Nestor PJ, and Rogers TT. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience,* 8(12): 976-987, 2007.

Pobric G, Jefferies E, and Lambon Ralph MA. Amodal semantic representations depend on both anterior temporal lobes: Evidence from repetitive transcranial magnetic stimulation. *Neuropsychologia,* 48: 1336-1342, 2010.

Rogers TT, Lambon Ralph MA, Garrard P, Bozeat S, McClelland JL, Hodges JR, and Patterson K. Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review,* 111(1): 205-235, 2004a.

Rogers TT, Lambon Ralph MA, Hodges JR, and Patterson K. Natural selection: The impact of semantic impairment on lexical and object decision. *Cognitive Neuropsychology,* 21(2-4): 331-352, 2004b.

Rogers TT and McClelland JL. *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press, 2004.

Rogers TT and Patterson K. Object categorization: Reversals and explanations of the basic-level advantage. *Journal of Experimental Psychology: General,* 136(3): 451-469, 2007.

Sajjadi SA, Patterson K, Tomek M, and Nestor PJ. Abnormalities of connected speech in semantic dementia vs. Alzheimer's disease. *Aphasiology,* 26(10): 1219-1237, 2012.

Schock J, Cortese MJ, and Khanna MM. Imageability estimates for 3,000 disyllabic words. *Behavior Research Methods*, in press.

Stadthagen-Gonzalez H and Davis CJ. The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods,* 38(4): 598-605, 2006.

Visser M and Lambon Ralph MA. Differential Contributions of Bilateral Ventral Anterior Temporal Lobe and Left Anterior Superior Temporal Gyrus to Semantic Processes. *Journal of Cognitive Neuroscience,* 23(10): 3121-3131, 2011.

Warrington EK. The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology,* 27: 635-657, 1975.

Wilson SM, Henry ML, Besbris M, Ogar JM, Dronkers NF, Jarrold W, Miller BL, and
Gorno-Tempini ML. Connected speech production in three variants of primary
progressive aphasia. *Brain,* 133: 2069-2088, 2010.

Woollams AM, Cooper-Pye E, Hodges JR, and Patterson K. Anomia: A doubly typical
signature of semantic dementia. *Neuropsychologia,* 46: 2503-2514, 2008.

Table 1: Background neuropsychological data

Test	SD patients									Normative data	
	DB	BC	JW	GH	MJ	PS	JC	mean	s.d.	mean	Cut-off
Age	65	60	76	59	71	59	60	64.3	6.8		
Sex	M	M	M	M	M	M	M				
ACE /100	81	70	66	63	59	61	47	63.9	10.4	92.7	89
MMSE /30	25	28	27	26	26	27	20	25.6	2.6	29	24
Digit span (backwards)	6	6	4	4	6	5	4	5.0	1.0	4.8	3
Rey figure copy /36	36	35	36	36	34	36	34	35.3	1.0	34.0	31
CCT pictures /64	56	39	47	75	32	45	48	48.9	13.8	59.1	51
Word-picture matching /64	63	51	55	55	49	48	50	53.0	5.2	63.8	62
Picture naming /64	58	41	30	42	36	23	14	34.9	14.3	62.3	59
Letter fluency (FAS)	36	41	38	22	12	21	6	25.1	13.5	41.1	17
Category fluency	19	10	21	20	15	6	4	13.6	6.9	50.7	30

ACE = Addenbrooke's Cognitive Examination (Mathuranath et al., 2000). MMSE = Mini Mental State Examination (Folstein et al., 1975). Rey copy taken from Osterrieth (1944). CCT = Camel and Cactus Test (Bozeat et al., 2000). Word-picture matching and picture naming taken from the Cambridge Semantic Battery (Bozeat et al., 2000). Category fluency comprised three categories: animals, fruit and birds. Normative cut-offs are the minimum scores considered normal in the test instructions or, if this was not specified, two standard deviations below the mean in healthy participants.

Table 2: Characteristics of the speech sample for each participant

|		SD patients							Controls								Mean (s.d.)	
		BC	DB	GH	JC	JW	MJ	PS	C1	C2	C5	C8	C9	C11	C14	C16	SD	Controls
Token counts	Grand total	13841	7326	3910	1930	4216	4403	4553	1587	7190	5091	3512	1917	3950	5664	5561	5740 (3906)	4309 (1935)
	Words with data	11180	5891	3126	1425	3464	3621	3620	1288	5902	4135	2870	1492	3293	4684	4551	4618 (3173)	3527 (1605)
	Words analysed	3705	1871	997	383	1048	954	1139	430	1915	1315	970	489	1050	1479	1388	1442 (1089)	1103 (503)
Type counts	Grand total	1296	1165	662	304	708	585	664	464	1295	944	738	484	835	920	922	769 (344)	825 (269)
	Words with data	901	807	479	245	514	456	506	361	871	675	548	374	616	680	666	558 (223)	599 (168)
	Words analysed	693	608	311	145	344	292	336	221	656	491	387	235	443	496	484	390 (192)	427 (144)
For analysed tokens	Mean log frequency	2.53	2.53	2.67	2.56	2.54	2.68	2.63	2.52	2.38	2.42	2.46	2.48	2.44	2.52	2.49	2.59 (0.07)	2.46* (0.05)
	Imageability	438	419	397	425	418	413	391	443	436	431	422	420	436	423	425	414 (16)	430*(8)
	Semantic diversity	1.90	1.91	1.95	1.86	1.88	1.94	1.92	1.87	1.84	1.86	1.85	1.89	1.89	1.90	1.87	1.91 (0.03)	1.87* (0.02)

Table lists word counts in terms of tokens and types for each participant's speech sample. Counts include the grand total of all words produced by the participant, the total number of words for which frequency, imageability and semantic diversity data were all available (words with data) and the total number of words included in the analysis (after excluding function words and other semantically impoverished words, as described in the text). The final section gives the mean frequency, imageability and semantic diversity values for words included in the analysis. * indicates a significant difference between patients and controls (t-test; $p < 0.05$).

Table 3: Psycholinguistic values for example words of varying specificity

Word	Semantic diversity	Log frequency	Imageability
Cathedral	1.39	1.28	599
Church	1.56	2.27	616
Building	1.92	2.25	578
Place	2.27	2.87	377

Imageability values were obtained from the MRC database.

Figure 1: Log production rates for each frequency and imageability band
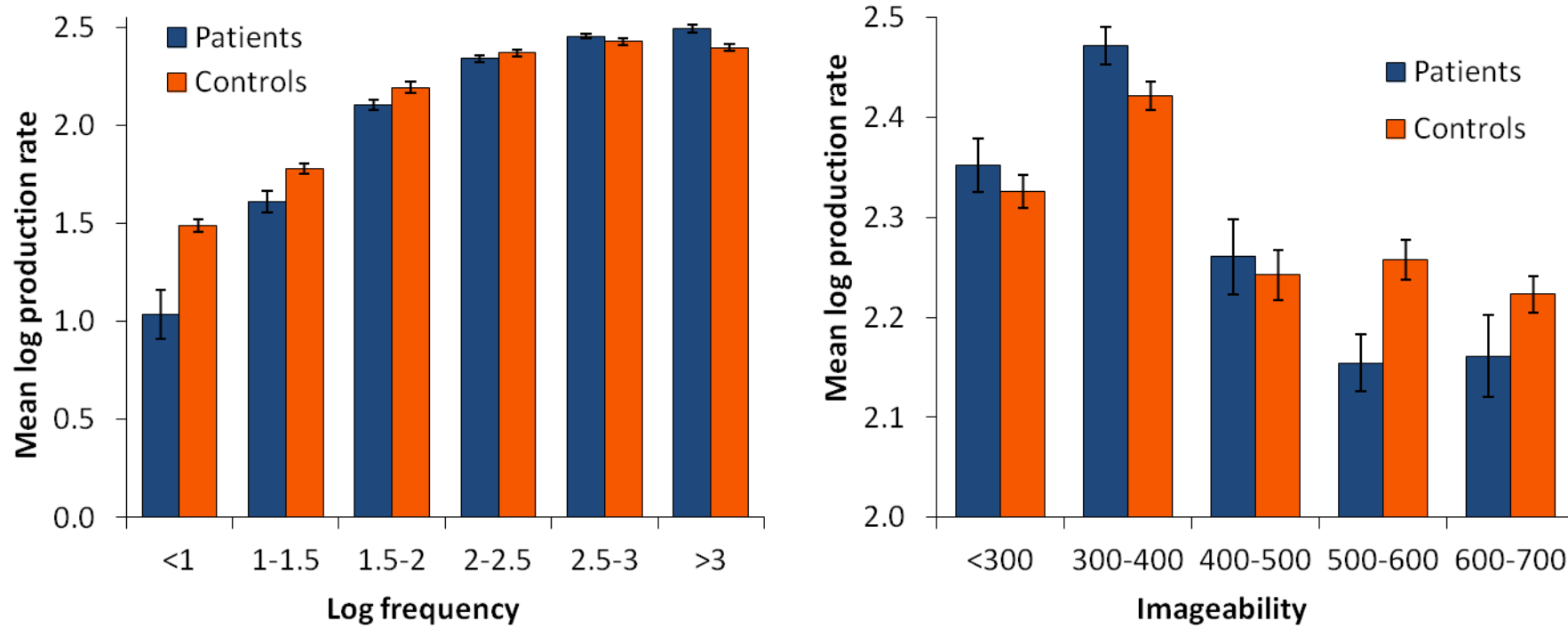
Error bars represent one standard error of mean.

Figure 2: Log production rates across the frequency-imageability space
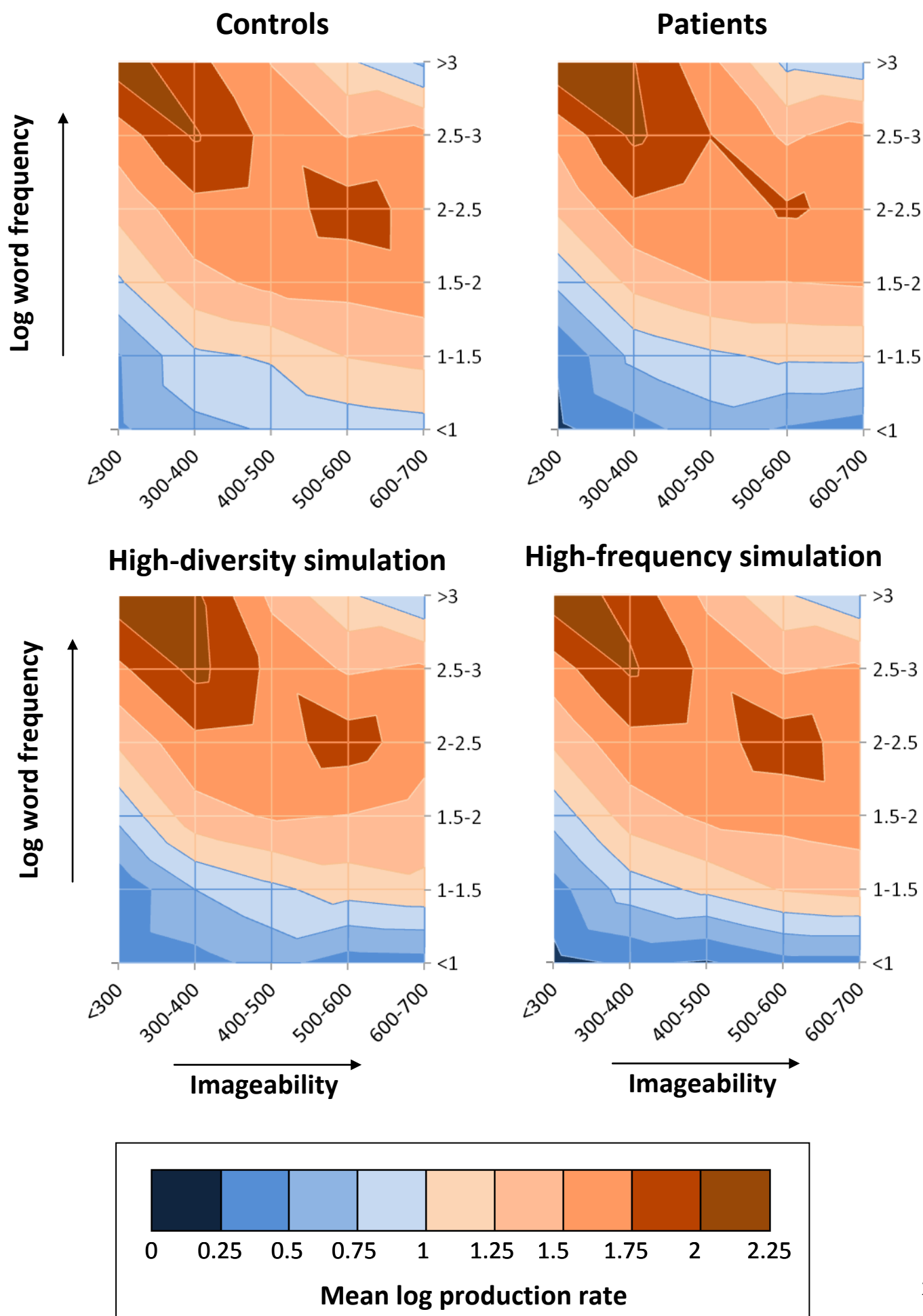
Figure 3: Plots of the difference values and p-values comparing patients with controls
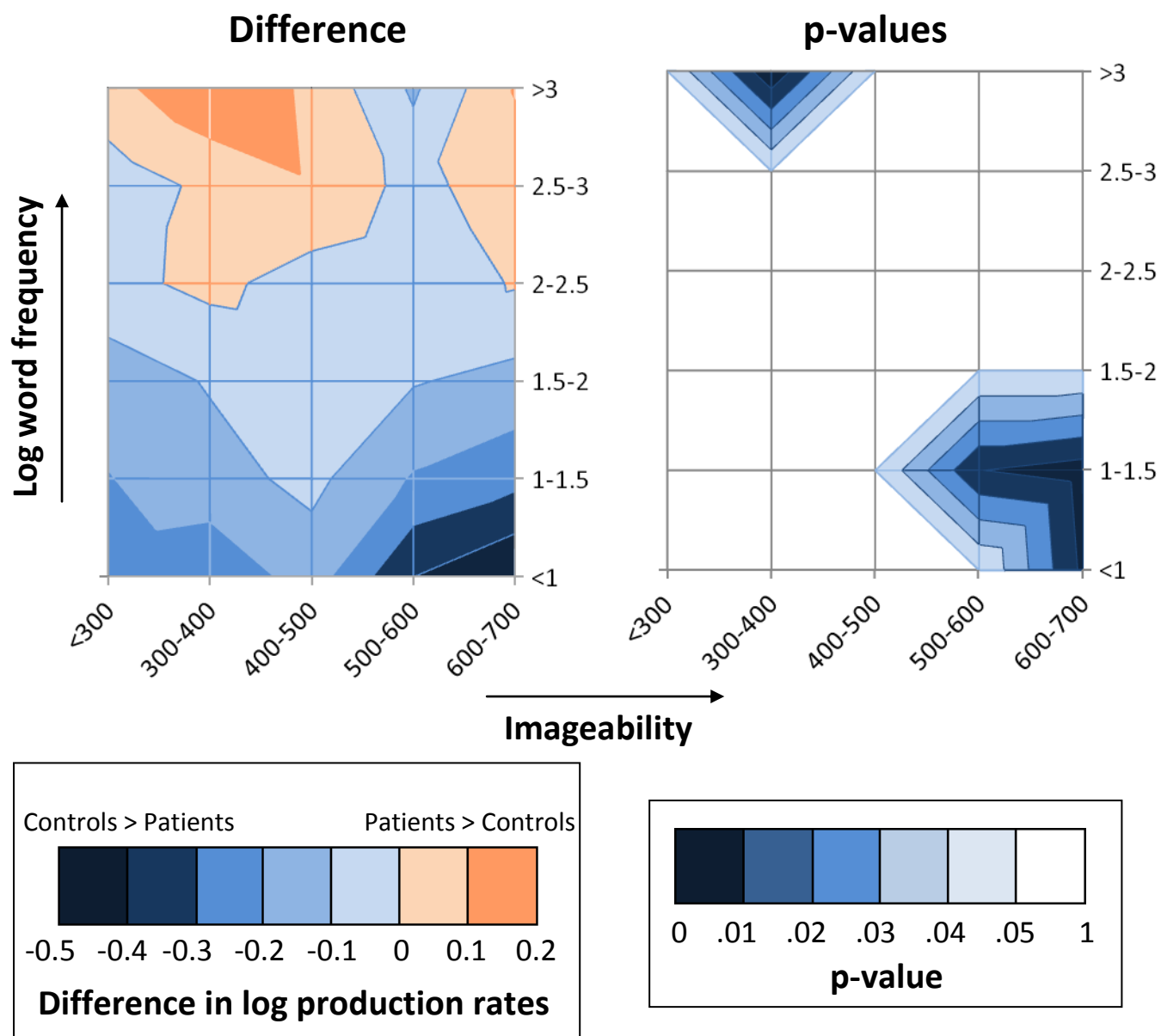
Figure 4: Mean semantic diversity of the words in each part of the frequency-imageability space
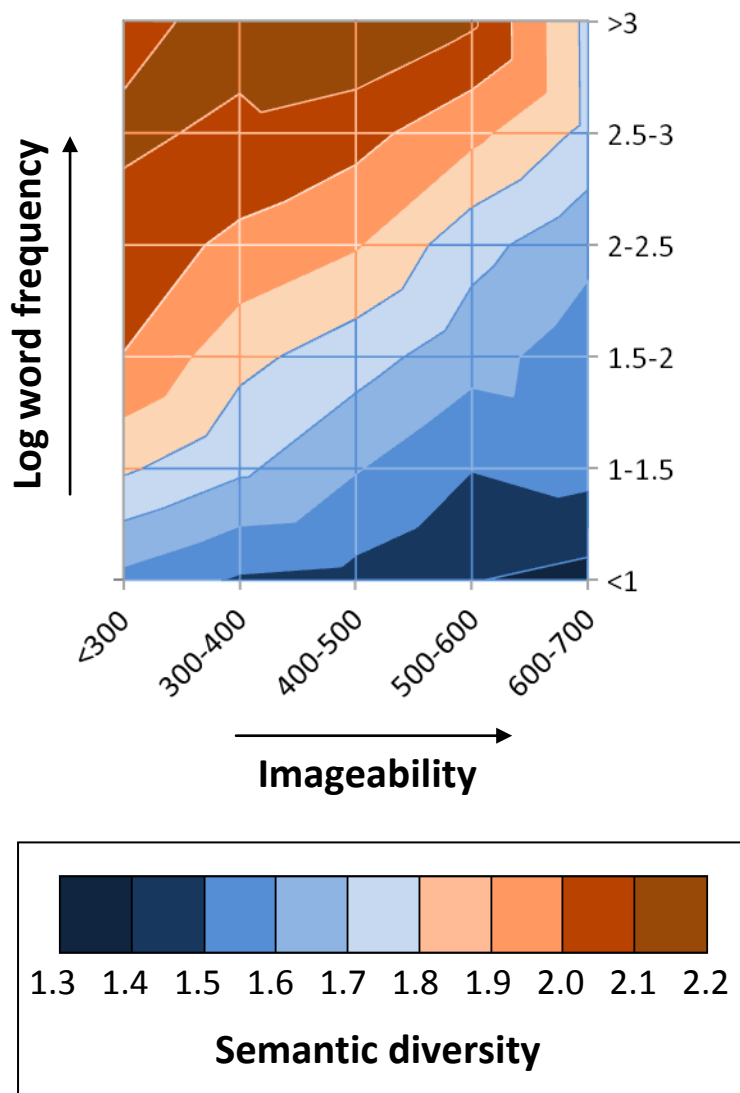
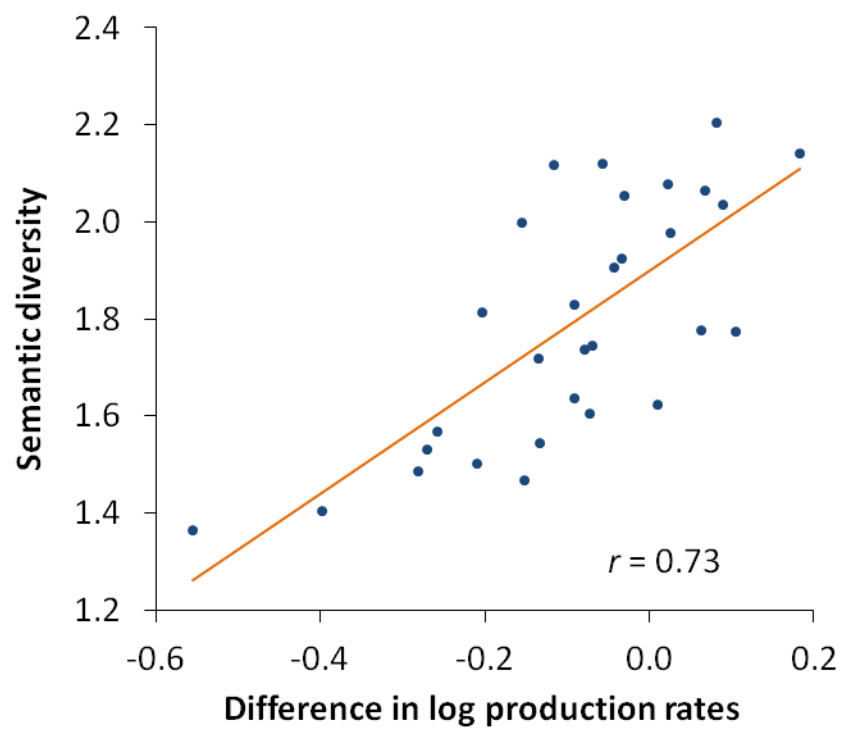Figure 5: Correlation between difference in word count and semantic diversity at each point on the frequency-imageability space
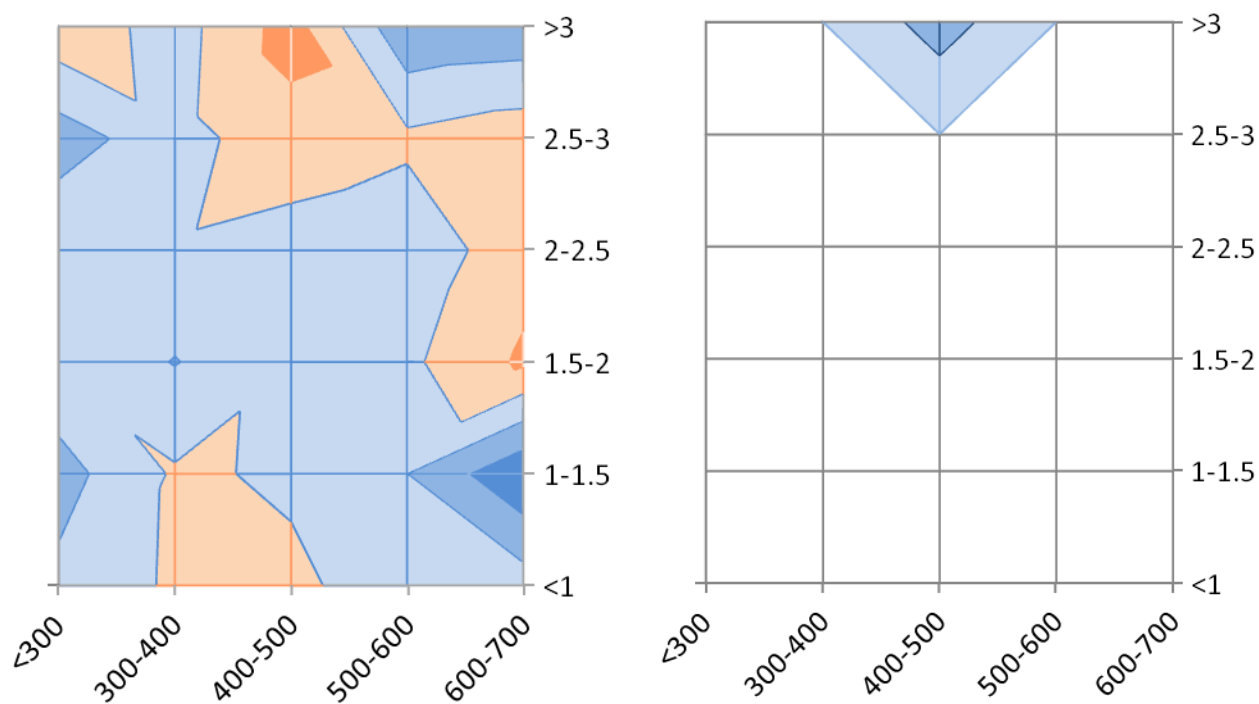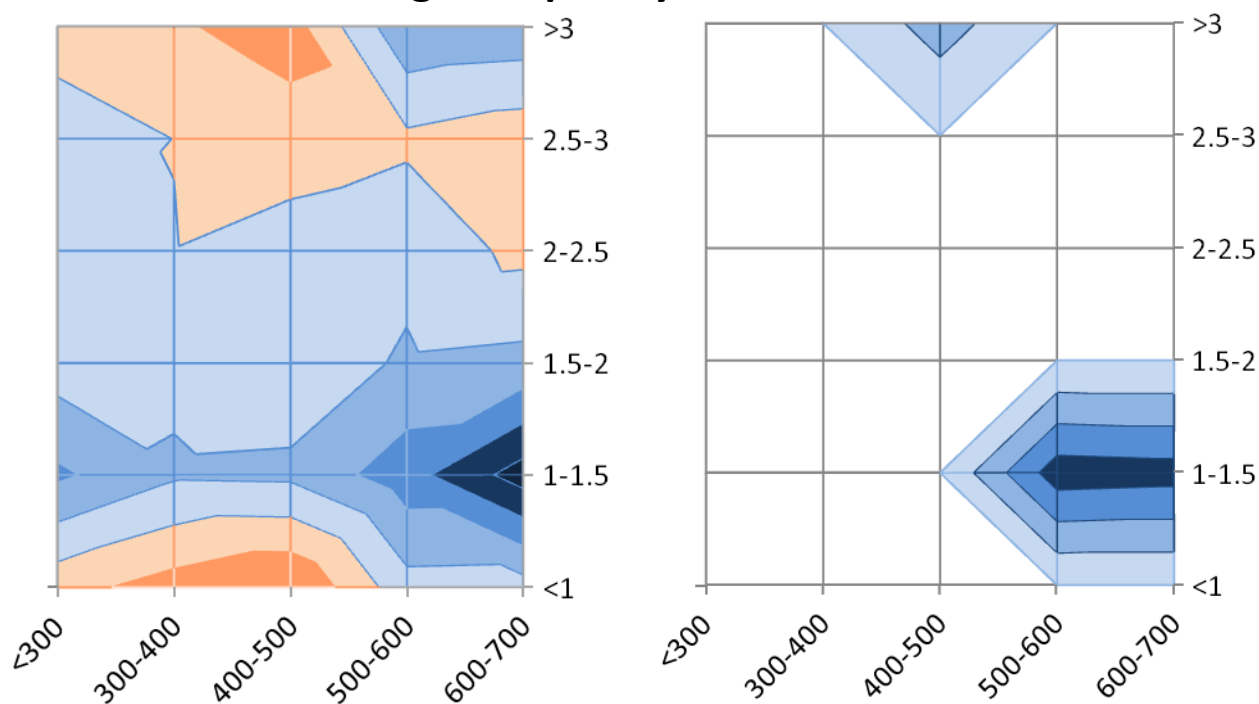
Figure 6: Plots of the difference values and p-values comparing patients with simulated data
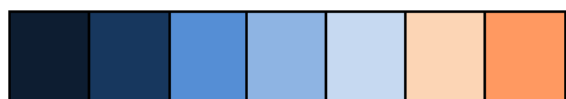
High-diversity simulation

High-frequency simulation

Simulation > Patients   Patients > Simulation

-0.5 -0.4 -0.3 -0.2 -0.1   0   0.1   0.2

Difference in log production rates

0   .01   .02   .03   .04   .05   1

p-value