Title:

A New Perspective on Word Association: how keystroke logging informs strength of word association

Michelle Aldridge[1], Lise Fontaine[1], Neil Bowen[2], Tamsin Smith[1]

[1]Cardiff University, School of English, Communication and Philosophy, Cardiff University.

[2]English Department, Thammasat University Klong Luang, Pathum Thani 12121, Thailand.

Abstract.

For many years, word association (WA) data has informed theories of the mental lexicon by analyzing the words elicited. However, findings are inconsistent and WA research is still waiting for 'a breakthrough in methodology which can unlock its undoubted potential' (Schmitt 2010: 248). In this paper, we offer a new perspective on WA by using keystroke logging (Inputlog, Leijten & Van Waes 2013) to captures the processes of word production. More specifically, we analyse pause behaviour during a continued, typed, word association task with 30 cue words eliciting 4 responses, per cue, to evaluate the strength of links in lexical selection processes. We show a strong positive correlation between pause length and inter-response location, providing empirical evidence which supports the established hypothesis that as more responses are elicited, links between them become weaker. Furthermore, using Fitzpatrick's response classification (2007), we found meaning-based responses were most common in the dataset generally, but, they particularly occurred after longer pauses, and exclusively so after the longest pauses. Position and form-based responses, whilst less frequent overall, typically followed the shortest pauses. In our conclusion we highlight the importance of our methodology in fine-tuning ongoing understanding of how we access the mental lexicon.

Keywords: Word associations, keystroke logging, pause behavior, lexical relations, mental lexicon.

1. Introduction

Within the field of lexical research, word association (WA) data are typically employed to further our understanding of how words are stored in our mental lexicon and to give insights into our semantic knowledge generally (Nelson, McEvoy, & Schreiber 2004). Though the research question may vary, the core of a WA task remains fundamentally the same: stimulus words (cues) are presented (orally or written) to a participant who is asked to respond with the first word(s) that come to mind. Such responses are seen to provide valuable information about an individual's 'concept building, the type of connections they maintain […] and how the strength of these connections develops and changes over time' (Zareva & Wolter, 2012:42). Thus, a lexical network metaphor (Meara, 1983; Wilks & Meara, 2002) has become a popular framework for illustrating that a word typically exists in an interconnected web of features and associations that both define and constrain its use (Aitchison, 2012). When a new word is encountered, it forms links to other nodes, where 'semantic link[s] are supplemented by phonological, orthographic, syntactic, and encyclopedic connections' (Wilks & Meara, 2002:303-304). In this view of word storage, the mental lexicon is seen as a highly-organized system of conceptual representations or 'nodes' (representing words), which are connected via links of varying strength. Accordingly, WA responses reflect the relative strength of the connections. Strength is often measured in terms of the number of people who respond with a particular word (DeDeyne, Navarro & Storms, 2013) or by the speed of a response, typically delivered orally (cf. Fitzpatrick 2007, Fitzpatrick & Izura 2011). As a consequence of this primary focus on word storage, much of the research captures the type or speed of the single response given (output) rather than the ongoing processes behind the response. Response classification has led to valuable insights into how our mental lexicons may be organized in terms of an elemental substratum of dominant connections: a network of frequent words connected by links of varying strength. However, once we move beyond the one-word response and look at multiple responses given to the same cue word, the associative links are less predictable and less well researched. We aim here, to fill this gap by adopting a novel methodology using InputLog (Leijten & Van Waes, 2013), a keystroke logging software package, to capture pause behavior during the production of multiple responses, enabling us to shed more light on how words are stored together.

Using keystroke logging software, we can examine the products of word associations (response types) and how pause behaviour changes over time (ordinal positioning of a response), which we argue reflects the strength of the association with the cue. By using response time (pause behaviour) during multiple WA responses, in this way, we gain insight about the relative strength of different lexical selection processes. The work presented here addresses two key research questions:

(i) Is there a correlation between inter-word pause length and the ordinal position of a response, and what might this suggest about the strength of associations?

(ii) Is there a correlation between inter-word pause length and the category of response selected, and what might this suggests about the strength of associations?

In what follows, we first review the key issues related to WA and the mental lexicon. Following this we consider the relevant theoretical background to word association, focusing, in particular, on pause behavior and word response times in language production, and what this can tell us about lexical processing. Section 3 provides the details of our methodology. Our results are presented in section 4, which are divided into two parts: First, we report on the average pause time between responses, and compare their place in the response sequence. Then, we analyse the categories of responses post pause and the lexical association to the cue word and to the immediately preceding response. In Section 5 we evaluate what our keystroke logging method contributes to findings in WA research. Section 6 summarises the main contributions of our research, and we briefly point to our plans for developing this research further.

2. Word association and the mental lexicon

Originally used in psychoanalysis, WA tasks have also been used over the years to investigate the link between language and the brain. As far back as Collins and Quillian (1969), research has shown that word associations are not arbitrarily organized but are linked through associated networks. However, there is no consensus on how association patterns might be interpreted. Nelson et al. (2000), for example, argued that responses in WA tasks reflect the strength of associations in memory, i.e. reading a cue word activates a set of related associates. The strongest associates in this set are then selected and reported. However, while there are some established patterns relating to semantics, syntax and phonology, there is considerable variation in responses, suggesting that participants are not homogeneous in their word selection patterns. For different participants, then, the same response can be represented at different strengths depending on their experience, age of acquisition, cultural background, or attention at any given moment (cf. Izura & Ellis 2002). For example, Entwisle (1966) demonstrated that children primarily produce clang (phonological form) and syntagmatic (collocational) responses, whereas, adults primarily produce paradigmatic responses (semantic). This change of associative behavior from clang to syntagmatic to paradigmatic chimes with what we know about the development of word knowledge and was the traditional way responses were classified. However, it is now accepted that this classification system is too crude (cf. Nation 1990) and more subtle systems are now often adopted. Fitzpatrick's (2007) system, for instance, categorizes associations as meaning-based, position-based, form-based or erratic, with each category further divided into 10 subcategories. This allows for a more fine-grained analysis of response types. Furthermore, much research suggests that a cue's word class impacts on WA test results: nouns, for example, elicit a higher proportion of nouns (paradigmatic responses) than verbs or adjectives. However, it must be noted that nouns are acquired early and, are, thus, well integrated and easily recognised, which may facilitate their elicitation by any type of cue word (cf. Bagger, Nissen and Henriksen's 2006).

Research considering the relationship between a cue word and the response(s) it elicits has also measured possible chaining effects (cf. Nelson et al, 2000), where a given response is related to the previous response rather than the cue word. It is the relationship between responses that is of particular interest to us in this study since this will allow us to infer the dominant types of lexical relations.

Turning now to response time, a common finding in WA research is that some words are retrieved and produced more quickly than others (cf. Plaut et al., 1996). Some have attributed this to 'age of acquisition' (AoA) (Izura & Ellis, 2002) and others to word frequency in the lexicon, but there is little agreement on the underlying mechanisms of these findings (cf. Balota et al., 2004). Concerning frequency, the most widely held assumption is that frequency is encoded as a weighting, where connections between units, whether they are words or abstract sub-lexical units (e.g., Plaut et al., 1996), are stronger when they are used together. Indeed, Beattie and Butterworth (1979:208) argue that pauses reflect the time taken to access less frequent lexical items; Navarrete et al., (2006) concur, noting that transitional probability and word frequency combined are the reasons behind hesitation pauses. In other words, associative links spread outwards from a cue word, activating any number of other words. This, also links to AoA, as early acquired words are likely to be more frequent in our lexicon. Brysbaert & Ghyselinck (2006) noted that an individual's AoA of a word can influence its selection in a WA task, with response times for early acquired words being typically 200ms plus quicker than late acquired words. This is arguably more advantageous when trying to map a complex network of (weak) associations based on relative strengths, such as those we would expect to see from multiple responses.

3. Method

In this study we use a novel method to examine the nature of multiple responses to cue words in a WA task. By considering both pause time between responses and the classification of type of response we are able to contribute a new perspective on lexical selection processes given that research to date has tended to focus on a single response.

3.1. *Participants*

Twenty-seven adult participants took part in the WA task. They were all native English speakers enrolled on a University undergraduate English Language degree course. Data collection was anonymous, and all participants gave formal consent for their participation. The relevant research proposal was authorized by the University's Ethics Committee.

3.2. *Materials*

The WA task used 30 cue words taken from Fitzpatrick and Clenton (2010), a tried-and-tested cue word set. These words are in the first thousand most frequent words in British English (cf. the British National Corpus, http://www.natcorp.ox.ac.uk), the set consists 19 nouns, 8 verbs and 3 adjectives.

3.3. *Procedure*

We chose a continued test mode where the stimulus word was displayed once and the participant was asked to give four responses resulting in a quantitative collection method to record inter-word pause length. Multiple replies were needed to measure strength of response, which gives us not only a more complete picture of participants' mental lexicons (Thomas 2007; Zavera 2007) but also an indication as to whether production time increases as more associates are produced. A continuous test method runs the risk that respondents will associate their responses to the last response given rather than the cue word. We address this issue of 'chaining', i.e. the response is activated by the previous response rather than the cue, in Section 5.3.

Data collection all took place on the same day. Participants were divided into 6 groups of 4/5 per group. Each participant was provided with an identical laptop on which to type their responses. Inputlog (Leijten & Van Waes, 2013) was installed on each laptop and was used to record their typing activity. Our interest was in the response time of multiple responses and, therefore, we recorded pauses between responses. With four responses to each cue word, we obtained 3 inter-response pause times: between responses 1 and 2 (1→2), 2 and 3 (2→3), and 3 and 4 (3→4).

Cue words were presented in a large font, one at a time, on a whiteboard in front of all the participants, (all of whom had equal opportunity to see each word). For each of the 30 cue words, participants were given 30 seconds to type 4 responses. 30 seconds was decided after running a pilot test with 5 participants which indicated this was the optimal time to elicit 4 responses. Thirty seconds was also decided upon to address the known finding that if participants are forced to respond too quickly, clang (similar sounding) responses are common (Kess 1992) while, if there is no time limitation many more pauses are likely to be idiosyncratic (Kess 1992) and/or reflect more distractions such as looking at a phone, having a drink, etc.

At each 30 second interval, the researcher said, 'new word', and the next word was flashed onto the screen. Participants recorded their responses by typing them into MS Word. The task onset was synchronised by having participants type 'Now we will begin' at the top of their Word document. Participants were instructed to: (i) press the space key between responses; (ii) use the enter key after completing a set of responses for each cue word; (iii) type their responses as quickly as possible so as to increase the accuracy of between response pause times – if a participant takes their time while typing, for example, there is the potential that it could be easier for them to think of the next word whilst they are still typing the current word. They were also asked not to make any changes to their responses. This included fixing typos; hence, we are reasonably confident that the pauses did not comprise monitoring and revising (cf. Section 2). The cue words were presented in the same order to all groups. After all the 30 cue words were presented, the task ended, and participants were asked to stop the Inputlog recording and to complete a self-reflection survey on how they thought they responded to the task (the results of which are not included in this paper).

3.4. *Data handling*

Keystroke logging is an effective means of capturing typed language production by recording all aspects of typing (e.g. keystrokes) as a time-stamped event. Inputlog was used here to record the pause durations between typed responses. However, the analysis of pause behavior is not standard in the literature, and, therefore, we needed to establish what we considered as a pause. For written language, the most used criterion is a threshold of 1second (Schilperoord, 2002) or more (Alves et al., 2007, 2012) for inactivity to be considered a pause for adults. However, some researchers, such as Olive and Kellogg (2002), adopt a threshold of 250ms on the grounds that this duration corresponds to the time needed to dot an 'i', or put an accent on a letter, or a bar across a 't'. This threshold is also supported by studies of typing for inter-key intervals (Grabowski, 2008; Aldridge & Fontaine, 2016; Aldridge & Fontaine, in press). Thus, we adopted the minimum 250ms threshold in this study.

Some pauses evidently mark 'scriptural inactivity' (Chenu et al 2014:1) but they may also reflect other activities such as cognitive, socio-physical, and physical behaviours (e.g., head scratching, adjusting a chair's position, looking out a window, etc. (Chenu et al., 2014). They may also be associated with anterior activity (e.g., fatigue, revision) or posterior activity (anticipation – what to write or do next). The main types of cognitive processes identified as correlates for pauses in writing are retrieving, planning, formulating, monitoring, and repairing (Flower and Hayes, 1981; Levelt, 1983). With our methodology, only retrieving, planning, and formulating should be occurring (cf. section 3.3) because participants were asked not to monitor their typing. The interpretation of a pause, thus, is multifaceted, and research suggests that we must take into account several factors, many of which are introduced in the following paragraphs and sections.

Pauses between responses were recorded for each participant (see Figure 1). We then divided the data into two equal sets to investigate the relationship between pause lengths (short or long) and response type. This resulted in a set of 15 responses with the shortest pause length and another set of 15 with the longest pause length. The decision to focus on the 15 shortest and 15 longest response times was a form of inductive content analysis, and it allowed us to identify emergent patterns from a very large dataset (a total 2414 responses were recorded). This data was collated into two tables comprising details of the participant, cue word number (1-30), place of pause, pause length, relation to cue word, and relation to previous response. Responses were categorized, using Fitzpatrick's (2007) method, as follows: meaning-based responses determined by semantic characteristics; position-based responses determined by syntactic and collocational characteristics; form-based responses determined by phonological, orthographical or morphological characteristics; and erratic responses, where no link between cue and response was apparent to the researcher, even though there was clearly some association for the participant.

4. Data analysis

To address our first research question, which concerns the relationship between pause length and pause location, the individual and mean pause lengths across the three between-pause locations were measured (pauses located between words 1→2, 2→3, and 3→4). Then, the 15 shortest and 15 longest pause lengths across all participants and locations were detailed with the emerging patterns between pause length and location. We then address our second research question, which concerns the relationship between inter-word pause length and response category, by examining the response categories and relations to cue words/previous responses. We go on to discuss any emerging relationship between the lexical relations of the response and the length of the pause preceding it, using Fitzpatrick's (2007) classification of responses: meaning-based, form-based, and position-based and erratic.

4.1. *Overall mean results of pause lengths*

Our first finding is that the average pause length differed significantly depending on where the pause occurred. More specifically, there was a positive correlation between pause location and mean pause lengths across participants, as shown in Table 1, 17 participants did not give a fourth response, and, thus, these pause times are not included in the mean results, but one may assume that they would have contributed to a further increase in the average pause length for the 3→4 location, given that they did not have sufficient time to access a fourth response.

TABLE 1: MEAN PAUSE LENGTH

Inter-response pause location	Mean pause length (ms)	Total responses
Between responses 1→2	1747.06	810
Between responses 2→3	2560.15	808
Between responses 3→4	2985.64	796

A comparison of pause length and pause location was made using a one-way ANOVA. This showed a statistically significant mean difference between the pause length for the 1→2 and 2→3 at the $p < .01$ level [F (1, 52) = 9.46, p =0.003)], where responses were, on average, 813.09ms slower at the 2→3 location. Similarly, there was a significant difference between the pause time for locations between 1→2 and 3→4 at the $p < 0.01$ level [F (1, 52) = 17.07, p = 0.0013)], where responses were, on average, 1238.59ms slower at 3→4 than at the 1→2 location. Between 2→3 and 3→4, however, there was no significant difference in mean values [F (1, 52) = 1.56, p = 0.21). Although this final difference is statistically non-significant ($p > 0.05$), it still contributes to an overall patterning between progressive pause locations and increasing pause durations; i.e. with each subsequent response, the average response time increases. In brief, these results show the pause time between words 1→2 to be the shortest (see Figure 1), and through continued response time recordings, we clearly show how the strength of the response weakens as it takes longer to give each additional response.

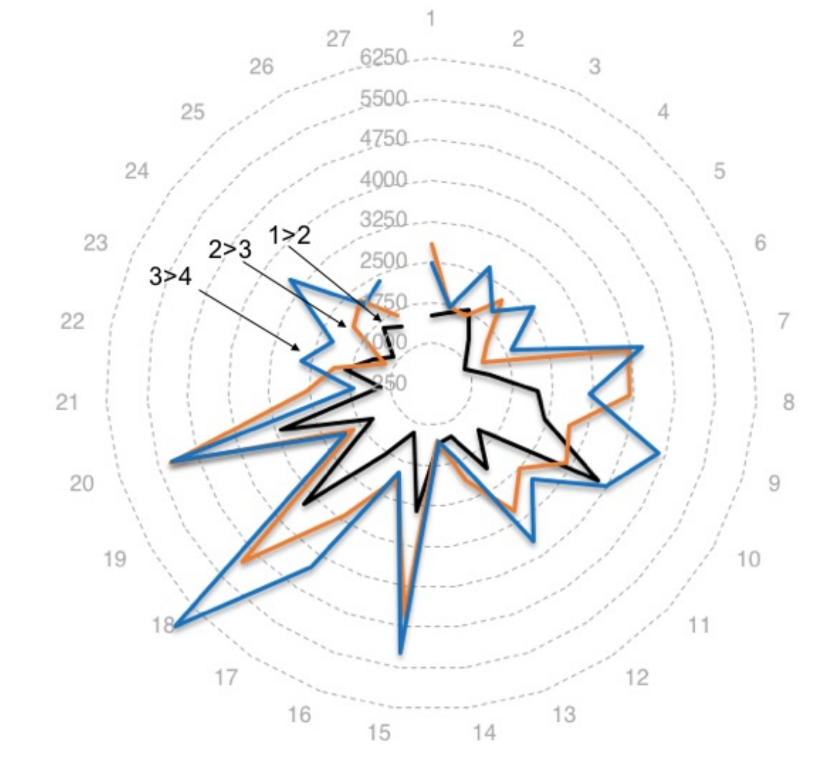

FIGURE 1: INDIVIDUAL AVERAGE PAUSE LENGTHS PER PARTICIPANT

4.2. *Response type*

To address the question of whether pause length impacted on response type, we chose to study the response following the 15 shortest (i.e. those above but closest to 250ms) and the 15 longest pauses (all above 13,634ms across all the participants and cue words). The 15 fastest responses amongst all participants are presented in Table 2.

TABLE 2: WORD RELATION CATEGORIZATION AMONG THE FIFTEEN SHORTEST PAUSE LENGTHS (250MS)

Participant	Cue	Pause location	Pause length	Relation to the cue word	Relation to the previous response
P42	26	1→2	250ms	position (consecutive yx collocation)	Similar form not meaning
P27	10	2→3	250ms	position (consecutive xy collocation)	erratic association
P1	5	1→2	250ms	form	erratic association
P1	6	1→2	250ms	meaning (synonym)	meaning (lexical set/context related)
P1	8	1→2	250ms	meaning (conceptual association)	meaning (conceptual association)
P1	25	1→2	250ms	meaning (lexical set/context related)	meaning (conceptual association)
P1	30	1→2	250ms	position (consecutive yx collocation)	meaning (conceptual association)
P30	2	1→2	250ms	meaning (conceptual association)	Erratic response
P30	18	3→4	250ms	position (consecutive xy collocation)	position (consecutive xy collocation OR Specific synonym*)

P14	22	3→4	250ms	erratic association	meaning (conceptual association)
P37	1	1→2	250ms	meaning (conceptual association)	meaning (other colloquial)
P18	1	2→3	250ms	meaning (lexical set/context related)	erratic association
P16	26	1→2	250ms	meaning (conceptual association)	position (consecutive yx collocation)
P38	21	2→3	250ms	form	meaning (conceptual association)
P26	9	1→2	250ms	position (consecutive xy collocation)	meaning (lexical set/context related)

From Table 2, when looking at only the relation to the cue word (4^{th} column), 6 of the quickest responses were meaning-based, 5 were position-based (xy/yx collocation), 2 were form-based, and 2 were erratic. Moreover, there is very little 'matching' in this set between the relation to the cue (4^{th} column) and the relation to the previous response (5^{th} column) – i.e., only 2 associative types were 'repeated' amongst the 15 quickest responses: a conceptual association, where response 1 to cue word 8 was the same type as response 2 for P1, and a consecutive xy collocation, where response 4 was the same type of response for both cue word 18 and response 3 for P30.

The 15 longest responses amongst all participants are given in Table 3. Here, we see that, 86.7% of relations in the longest pause category have meaning-based associations to the cue word or previous response, and there are no xy collocations at all. Moreover, when we compare the responses following the longer pauses there are 4 occasions where relations to a cue word and a previous word are the same. Clearly, this is a small sample, but it suggests that the more the mental lexicon is accessed for word selection, the more there is some kind of convergence, in the sense that word n and n-1 reinforces or strengthens the search.

TABLE 3: WORD RELATION CATEGORIZATION AMONG THE FIFTEEN LONGEST PAUSE LENGTHS

Participant	Cue word	Pause location	Pause (ms)	Relation to cue word	Relation to previous response
P33	15	3→4	33431	meaning (synonym)	erratic association
P42	21	3→4	26957	meaning (conceptual association)	meaning (lexical set/context related)
P22	20	1→2	26395	meaning (synonym)	meaning (lexical set/context related)
P29	1	3→4	22760	meaning (conceptual association)	erratic association
P33	24	3→4	21279	meaning (conceptual association)	meaning (conceptual association)
P37	25	2→3	19687	meaning (conceptual association)	erratic association
P22	27	1→2	17332	meaning (lexical set/context related)	meaning (synonym)
P33	3	3→4	16427	meaning (conceptual association)	meaning (conceptual association)

P19	3	3→4	15818	meaning (conceptual association)	position (consecutive xy collocation)
P33	7	3→4	15585	meaning (lexical set/context related)	meaning (conceptual association)
P33	23	2→3	15570	Specific Synonym	meaning (lexical set/context related)
P20	23	3→4	14867	meaning (conceptual association)	meaning (conceptual association)
P19	28	2→3	14368	meaning (lexical set/context related)	meaning (lexical set/context related)
P22	13	3→4	14196	meaning (lexical set/context related)	meaning (lexical set/context related)
P33	30	2→3	13634	meaning (lexical set/context related)	erratic association

A clear trend is that meaning-based responses are by far the most common across both the shortest and the longest pause lengths. Interestingly, when position and form-based associations occur, they are produced only following short pauses. The lack of any position or form-based associations within the longest pause group is a very interesting finding, which we discuss below.

5. General discussion

This section will first look at the average duration between word pauses, before looking at the categories of associations and possible reasons as to why meaning-based responses were consistently chosen over other types of responses in the 15 longest pauses category. We will also briefly consider ‘chaining’, and the effect this may have had on response categories. In addition, we offer our evaluation of using keystroke logging as a methodology for accessing the strength of word association responses.

5.1. *Average duration of between word pauses*

The results showed a strong positive correlation between pause location and pause length, with between -word pause duration lengthening as the number of responses to a cue increased. Consequently, we saw longer average pause times between 3→4 than 2→3 and 1→2, suggesting that as the number of responses increases, more time is needed to select an association. Thus, the strength between association 1 is strong but becomes weaker with each word accessed. It is worth noting here that this trend was not influenced by any particular cue word, as, for example, cue word 1 elicited two of the fastest responses, but, also, one of the slowest, whilst cue word 21 elicited both one of the fifteen fastest and one of the fifteen longest pause times.

This pause length pattern suggests that we effectively ‘use up' our strongest (most ‘dominant’) associations quickly. It may, also, be the case that as we are required to produce more and more responses, the number of nodes that are activated in our mental lexicon increases. This spreading activation, whether from latent or active nodes, generates a wider network of items from which we can choose. For example, a cue word may activate links to other nodes based on their meaning, form, or position-based associations; the chosen response (response 1) then activates links to other nodes based on their meaning, form, or position-based associations. This occurs in a very short period of time, and it could be argued that the spreading activation of latent nodes (from the cue) and active nodes (from the response), whilst possibly overlapping somewhat, create a denser network of associative links from which we must then choose. This density continues to grow as we add further responses to the already selected ones, up to the point that the network of links is so densely populated that 17 of the 27 respondents (66.7%) failed to give a fourth response in time; at least once.

However, the findings also show that the average pause duration does not increase by an equal amount across the three pause locations. There is a significant difference of - 819.022ms ($p < 0.001$) between 1→2 and 2→3, and only - 404.881ms ($p = 0.009$) between 3→4. The fact that the pause duration increases most significantly in the 2→3 location may hold clues as to how the mental lexicon is organized. Although pause length continues to increase, the much smaller increase in duration between 2→3 and 3→4 suggests that associations should be thought of as lying on a continuum of strength, not equi-distance. If this pattern was to continue, one would expect to see the pause durations continue to increase, but the difference in duration between locations to become smaller and smaller, suggesting retrieval strategies may become more similar as the number of responses increase. This hypothesis corresponds with that put forth in the previous paragraph in relation to increasing densities of nodes from which to choose. However, it also suggests that the number of responses given may effectively saturate the number of nodes activated. After all, there is a limit to the amount of words a person knows, and this puts limitations on the number of associates they can recall in response to a set of words, particularly if that person is under the belief that each response must have some link to the original cue. We may also conclude that the ability to access links based on form association alone begins to lessen, and the significant pause increase may be where we become more dependent on meaning-based links. It would be interesting to see if this patterning of decreasing pause time differences would continue if the same participants were asked to give 6 or 8 responses, although, of course, we would also expect to see the number of non-responses increasing as a factor of an individual's level of word knowledge.

5.2. *Pause duration and response category: meaning, form and position-based associations*

We found a correlation between inter-word pause duration and the category of response selected. More specifically, the 15 longest pauses represented either meaning-based associations (86.7%) or erratic responses (13.3%). Meaning-based associations include synonyms, members of a lexical set, and conceptual associations (see Fitzpatrick (2007) for details). These findings suggest that when participants take longer to think of a word, they draw upon semantic connections, rather than syntactical or phonological connections. It is difficult to determine, from these results alone, whether we first search for position/form-based associations and if we are unable to find any quickly, we look for meaning-based associates. An analysis of 1st responses to the cue word may be revealing here, but this was not part of our research design since we were interested in pauses between responses (and there is no accurate measure of pauses between the cue and first response in our design). However, the large amount of meaning-based responses (50%) in the short pause group would suggest that meaning-based responses are often the first lexical item participants access; this echoes the work of Fitzpatrick and colleagues (2011), who recorded no form-based responses in their test with L1 speakers where only one response was given.

As already noted, meaning-based associations were the most prominent type of response amongst the 15 shortest pauses (50%). This was followed by position-based associations (26.7%), and then form-based associations (10%). As per the 15 longest pause times, 13.3% of the 15 shortest pause responses were erratic responses. The fact that 68.3% of the 15 shortest and 15 longest responses are meaning-based associates supports Entwisle's (1966) original finding, that adults show a predominance of paradigmatic (meaning-based) responses. Although verbs have previously been shown to produce the highest rate of syntagmatic (positional) responses (Cronin, 2002), supporting the view that they are most intimately related to syntax in their function (Pinker, 1994), only 3 of the 11 verbs listed in the 15 shortest or 15 longest pause groups elicited position or form-based responses. There were a similar number of verbs in both the shortest and longest groups, with 5 in the shortest and 6 in the longest group, suggesting that this word class bears little consequence to the selection of a position, form, or meaning-based response. However, where position and form-based associations occur, they are accessed very quickly; this is predictable given that form and position-based associations are finite. There are, therefore, fewer competitors to sort through when choosing an association which makes the response time quicker.

Another explanation for quick form-based response times for position and form based associations is that, because as children we produce a predominance of clang (form) and syntagmatic (position) responses, these items become firmly integrated into the adult linguistic system, and as such they may still be accessed very quickly. The relatively high percentage of form and position-based associations (36.7%) in the 15 shortest pause group indicates that certain words may elicit form or position-based associations, in line with research suggesting that AOA has an important effect on retrieval time (Morrison et al., 2002; Brysbaert & Ghyselincke, 2006).

In contrast, there are many meaning-based associations, and a slow retrieval could signify the time it takes to sort through possible competitors, or perhaps to select a word deemed appropriate; it may be that participants select a word but reject it, thinking it is not an accurate enough fit. Keystroke logging cannot capture this difference, but it is a question that could be asked in a post-task questionnaire.

5.3. *The effect of chaining on pause duration*

Fitzpatrick (2007: 322-323) notes that a response in a WA task is simply ‘a product of tension between two influencing entities: the cue word and the respondent'; however, research using WA tasks with multiple responses suggests that associations should also take into account the relationship to the cue word and any previous response(s) given (De Dayne et al., 2013). Our study found that previous responses seemed to influence participants' word selection in that connections could be drawn between them, but the response-response production here was not as prominent as these theorists suggest. Our results are more in keeping with Precosky (2011) who noted only a 1.3 % response chain in multiple responses. In our data, across the 15 shortest and 15 longest pause durations studied, the responses were always linked to either the cue word or a previous response. There were a total of 7 out of 30 instances with no linguistically identifiable link between the response and the previous response, and only 1 out of 30 where there was no link to the cue word. While the effect of a previous response is unknown, since we could not compare it to a control condition with no previous response, the fact that 22 out of 30 responses (73.3%) related both to the cue word and the previous response suggests that previous responses are used in combination with the cue word to select a new response. For example, P42 responded to the cue word *attack* with: art {250} heart {3822} violence {1591} knife (values given in braces indicate pause times in ms).

It could be argued that *heart* has been selected because it rhymes with *art* (similar form not meaning); however, 'art' could equally have been chosen due to its consecutive yx collocation with the cue word *attack* – the popular children's TV program (Art Attack) that aired on British TV 1990-2007.

As a response to these findings, 'chaining' seems an inappropriate term, as it suggests a chain of connections where one response can only be influenced by the word which directly precedes it. The pause time is instead affected by all previous responses, which can be thought of as one item in a set or local network of lexical items, which are activated and then used in combination with the cue word to select a new response. In brief, where an item is a strong 'hub' there is greater potential for multiple links (see example 1).

5.4. *Evaluation of keystroke methodology in word association*

Our results have shown two clear trends, namely (i) that pause length increases as more responses are elicited, suggesting increasingly weak links the more words are sought; and (ii) that longer

longer pauses are associated with meaning-based responses. Such quantitative data provides a baseline from which to determine further testable hypotheses. Such future work is needed to define the role of more qualitative issues; for example, the role of individual participant differences and the challenge of classifying responses (i.e. ensuring reliable classifications). We noted, for example, that there were 15 pauses that recorded the minimum 250ms; however, 5 of these were from the same participant. Similarly, the 15 longest pauses came from just 7 participants (n=27). Our results are, thus, in part the consequence of a few participants who were particularly fast or slow across all responses, and it cannot be known, with certainty, whether the lexical relations are due to (i) individual selection strategies, for example, P1 took 250ms to find an association which was similar in form not meaning to the cue word, and had no decipherable link to the previous response given, while P33 took 33431ms to type one association where the response was a defining synonym of the cue word, and thus very closely linked. Further (ii) they may signify a broader pattern of adult lexical storage and access; or (iii) may simply be the consequences of typing ability (it is interesting to note that only 2 participants recorded both 1 of the shortest and 1 of the longest pause times).

Lexical classification presents challenges that are well-known to researchers (cf. Fitzpatrick and Izura 2011) and we had similar difficulties. For example, some of our responses have dual-link associations. For instance, example (1) shows how participant P38 responds to the cue word *pot*:

(1) shot {281} food {250} hot (1341} not.

When looking at its relation to the cue word, *hot* may have been chosen because it rhymes with *pot* (similar form not meaning), or because of its conceptual association (pots typically get hot), or because of a consecutive yx collocation (*hot pot*). This is an example of how the relation may fall under all three categories of form, meaning, and position-based association. We may assume that because the first response *shot* rhymes with the last response *not*, *hot* is also selected because of its similar form relation with *pot,* rather than any relation with *food*, but this is by no means a certainty. In fact, it is likely that *hot* has been selected so quickly due to a combination of its many associative links with the cue word and the responses preceding it (i.e., it also rhymes with *shot*, has an overlap in orthography, and has a conceptual relation with food). To take another example

(2), participant P30 responds to the cue *hold* as follows:

(2) tight {250} together.

The pair *hold* and *tight* appears to be a consecutive xy collocation, as does the pair *hold* and *together*. By this same logic, *hold* and *tight* and *together* could also be a consecutive xyz collocation of three words, demonstrating 3 position-based responses based on frequently occurring collocations. However, the words *tight* and *together,* in colloquial slang, could also be used as specific synonyms (and therefore meaning-based associations may have been chosen due to similarity and not frequency). These form and position-based associates demonstrate the difficulty when classifying a response. In many of these cases, the form or position-based response could also be related in meaning in some way, which demonstrates that responses can be in more than one category, as it is impossible to know which association the participant was using when selecting their response. Indeed, a combination of similarity and frequency, and meaning and position or form-based associations may have been used. Previous research has shown that such dual-link associations are stronger and quicker to access (Fitzpatrick & Izura, 2011) but very infrequent (cf. Fitzpatrick et al., 2013). However, our data suggest they may in fact be much more common, although this needs to be pursued in future research.

6. Conclusion

Our innovative use of keystroke logging has furthered our understanding of how pausing in WA tasks can reveal information about how words are stored and their relative strength to one-another when accessed. More specifically, through multi-response analyses, this study has found evidence of both (i) a strong positive correlation between inter-word pause length and the place in which the pause occurs, and (ii) inter-word pause length and the category of response selected.

The finding that all 15 of the longest pause durations (excluding erratic responses) elicited meaning-based responses would suggest that form or collocational associations are lost quickly. It seems that in auto activation we access form, but as soon as we enter lexical 'search' mode we access meaning-based associations. However, 50% of the quickest responses were also meaning-based associations, demonstrating that meaning-based associations are the dominant type in a typical adult's L1 mental lexicon.

As discussed, the fact that position and form-based associates were only found in the 15 shortest pause times could be due to a combination of two reasons. Firstly, the early acquired clang and syntagmatic associates used in childhood (see Entwisle, 1966; Cronin, 2002) are so deeply integrated into the mental lexicon that these words are still relied upon and retrieved most quickly by adults when in high pressure situations. The second possible reason is that these words are retrieved quicker due to the fewer position or form-based competitors one would have to sort through, in comparison to the many meaning-based associations adults store in the mental lexicon. This is only plausible, of course, if position, form, and meaning-based lexical items are not all accessed together – if this were the case one would expect these findings to be due to the first reason only. The small analysis into the use of concrete nouns introduced above seems to suggest that this is the most likely reason.

Moreover, this study suggests that adults select form or position-based associations very quickly and often do so before meaning-based associations. However, common sense would dictate that in adult language, form-based associations are less likely to be selected because words which rhyme but have no related meaning are less frequent in every day speech than words which share the same semantic field (meaning-based associations). Consequently, our results show that it is not just the frequency of a word that effects word selection. By looking at pause duration, we can see that when words are selected quickly rather than slowly, this selection is often based on form or position-based similarity, or is based on some other factor, such as the effects of AoA and the long-term integration of a word and its associates into the mental lexicon.

Given the scarcity of empirical research into the mechanics of responses in WA tasks, attention must be given to replicating these results and to testing the claims we have made using different methods. A wider, more heterogeneous participant set is also needed. For example, there were 15 pause times at the chosen minimum of 250ms, yet 5 of these were from the same participant. We also need to control the influence of individual performances more carefully– it may be that some of the pause differences are because some participants are better typists, and/or more confident individuals, or simply more motivated to answer, etc. It is important going forward that we control individual profiles to ensure that we are tapping into processes. One way to achieve this is by replicating work by Van Waes et al 2017) in capturing personal characteristics through copy tasks recorded by Inputlog. Furthermore, we would want to determine whether the lexical relations identified are due to individual selection strategies or signify how the adult mental lexicon is organised. Similarly, by increasing our understanding of how the mental lexicon is organized in a range of demographic groups, this type of WA methodology, using fine-grained recordings of typed response times, may also then lend itself to neuropsychological testing, such as tests designed to aid in the diagnosis of degenerative brain diseases like Alzheimer's (Leijten et al., 2014). Moreover, while we did not test the effect of the word class of the cue words (e.g. concrete verbs vs other types of verbs vs. nouns etc.), we predict that such research will be very fruitful, allowing us to conclude whether word class impacts on the type of response given (e.g. position or form-based responses).

Our results concentrate on multiple responses, it would now be useful to repeat this study by capturing the production time of the first response in relation to the cue word. This is a more complex measurement and not entirely comparable to between response pauses, but nevertheless it is worth pursuing in order to see whether the same results appear. For example, if a long pause occurs after the cue word prompt, do we find only meaning-based responses as we did in our results (cf. Table 3)? To measure the response time between the cue and the first response, we would have to alter our research design to control for: (a) the exact moment the participant is exposed to the cue, and (b) their hand placement in relation to the keyboard at the time they are presented with the cue. Although these two factors could theoretically be brought into alignment, typing proficiency could still arguably affect the speed of response. Therefore, perhaps some form of quantile normalization could also be introduced in further studies, where a pre-test control task could be used to measure typing proficiency. The results of these measures can then be subsequently used to factor out differences in typing proficiency.

We believe that this paper sets out an important research agenda because our keylogging methodology evidenced that long inter-response pauses produced only meaning-based responses. This result needs to be explored further, and indeed our plan is to replicate this work in future studies and to extend it in to the areas discussed. To conclude, by using Inputlog, we have been able to highlight the overall contrast in pause duration for position, form, and meaning-based associations more clearly than indicated in research elsewhere.

REFERENCES

Aitchison, J. (2012) *Words in the Mind: An introduction to the mental lexicon.* (4th Ed.) Chichester: Wiley.

Aldridge, M. & Fontaine, L. 2016. The impact of mode on writing processes: A cognitive functional perspective on student writing. *Special Issue on SFL and Development in Writing and Reading Competence in Language Education, Revista de Lenguas y Culturas Extranjeras*, 7, 7-32.

Aldridge, M. & Fontaine, L. In press. The role of working memory in typing fluency: a new perspective on writing activity In Lindgreen, E., Knospe, Y. & K. Sullivan *Researching writing with observational logging tools from 2006 to the present and beyond.*

Alves, R. A., Castro, S. L., de Sousa, L. and Strömqvist, S. 2007. Influence of typing skills on pause-execution cycles in written composition. In M. Torrance, L. van Waes and D. Galbraith (eds.) *Writing and cognition: Research and applications.* Amsterdam: Elsevier, 55–65.

Alves, R.A., Branco, M., Castro, S.L. & Olive, T. 2012. Effects of handwriting skill, output modes, and gender on fourth graders' pauses, language bursts, fluency and quality. In V. W. Berninger (Ed.)*, Past, present, and future contributions of cognitive writing research to cognitive psychology*. New York, NY: Psychology Press, 389-402.

Bagger Nissen, H. & Henriksen, B. 2006. Word class influence on Word association test results. *International Journal of Applied Linguistics* 16(3): 389-408.

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.

Beattie, G. & Butterworth, B. 1979. Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and* Speech 22(3): 201-210.

Brysbaert, M. & Ghyselinck, M. 2006. The effect of age on acquisition: Partly frequency related, partly frequency independent. *Visual Cognition* 13 (7/8): 992-1011.

Chenu F., Pellegrino F., Jisa H. & Fayol M. 2014. Interword and intraword pause threshold in writing. Front. Psychol. 5:182 10.3389/fpsyg.2014.00182.

Collins, A. M. & Quillian, M. R. 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 241-248.

Cronin, W. 2002. The syntagmatic-paradigmatic shift and reading development. *Journal of Child Language* 29(1):189-204

DeDeyne, S., Navarro, D. & Storms, G. 2013. Better explanations of lexical and semantic cognition using networks derived from continued rather than single-word associations. *Behavioral Research methods*. 45 (2) 480-98

Entwisle, D. R. 1966. *Word associations of young children*. Baltimore: John Hopkins Press.

Fitzpatrick, T. 2007. Word association patterns: unpacking the assumptions. *International Journal of Applied Linguistics* 17(3):319-331.

Fitzpatrick, T. & Clenton, J. 2010. The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, 27(4): 537-554

Fitzpatrick, T. & Izura, C. 2011. Word association in L1 and L2: An exploratory study of response types, response times and inter-language mediation. *Studies in Second Language Acquisition*, 33:373-98.

Fitzpatrick, T, Playfoot, D, Wray, A. & Wright, M. 2013. Establishing the Reliability of Word Association Data for Investigating Individual and Group Differences. *Applied Linguistics* 203: 1-29.

Flower, L. & Hayes, J. 1981. *A cognitive process theory of writing. College Composition and Communication*, 32, 365-387.

Grabowski, J. 2008. The internal structure of university students' keyboard skills. *Journal of Writing Research* 1(1): 27-52.

Kess, J. F. 1992. *Psycholinguistics: Psychology, linguistics and the study of natural language.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Izura, C. & Ellis, A. 2002. Age of acquisition effects in word recognition and production in first and second languages *Psicológica*, 23, 245-281.

Leijten, M. & Van Waes, L. 2013. Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. **Written Communication** 30(3), 358–392 [available online: www.inputlog.net/description.html Accessed: 314/03/2017).

Leijten, M., Engelborghs, S., Van der Mussele, S., Van Horenbeeck, E., Mariën, P. & Van Waes, L. 2014. *Cognitive writing process characteristics in Alzheimer's disease.* Paper presented at the Conference on Writing Research, Amsterdam.

Levelt, W. 1983. Monitoring and self-repair in speech. *Cognition.* 14 (1) 41-104.

Meara P. 1983. Word associations in a foreign language: a report on the Birkbeck vocabulary project, *Nottingham Linguistic Circular*, vol. 11 (2):29-38

Morrison, C.M., Hirsh, K.W., Chappell, T. & Ellis, A.W. 2002. Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *European Journal of Cognitive Psychology* 14(4): 435-459.

Nation, I.S.P. 1990. *Teaching and Learning Vocabulary* New York: Newbury House.

Navarrete, E., Basagni B., Alario F. X. & Costa A. 2006. Does word frequency affect lexical selection in speech production? *The Quarterly Journal of Experimental Psychology* 59(10):1681-1690.

Nelson, D, L. McEvoy C, L. & Dennis, S. 2000. What is free association and what does it measure? *Memory* & Cognition 28(6):887-899.

Nelson, D., McEvoy, C. & Schreiber, T. 2004. The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, 36, 402–407.

Olive, T. & Kellogg, R. 2002. Concurrent activation of high and low level production processes in written composition. *Memory & Cognition*, 30: 4. 594-600.

Pinker, S. 1994. *The language instinct*. New York: W. Morrow and Co.

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.

Precosky, K. 2011. *Exploring the mental lexicon using word association tests: How do native and non-native speakers of English arrange words in the mind?* University of Birmingham (unpublished PhD).

Schilperoord, J. 2002. "On the cognitive status of pauses in discourse production," in *Contemporary Tools and Techniques for Studying Writing;* Olive T., Levy M. C., editors. (Dordrecht: Kluwer Academic Publishers) 61–87 10.1007/978-94-010-0468-8_4.

Schmitt, N. 2010. *Researching vocabulary: a vocabulary research manual* Palgrave Macmillan.

Thomas, D. 2007. Second language word associations: How Japanese make mental links between English words they have learned. *Unpublished assignment. University of Birmingham*.

Van Waes, L., Leijten, M., Marien, P. & Engelborghs, S. 2017. Alzheimer's disease: An exploration of copy tasks *Computers in Human Behavior*, 73, 311-319.

Wilks, C. & Meara, P. 2002. Untangling word webs: graph theory and the notion of density in second language word association networks. *Second Language Research*. 18.4: 303–24.

Zareva, A. & Wolter, B. 2012. The ‘promise’ of three methods of word association analysis to L2 lexical research. *Second Language Research*, 28 (1) 41-67

Zareva, A. 2007. Structure of the second language mental lexicon: How does it compare to native speakers' lexical organization. *Second Language Research* 23.2, 123-153.