Reading Aloud in the Context of the Task Set Paradigm: New Perspectives

by

Shannon O'Malley

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Psychology

Waterloo, Ontario, Canada, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In a series of five experiments I examined whether intention (as operationalized by task set) affects the processes involved in reading. The Task Set paradigm (Besner & Care, 2003) was used in all experiments. On each trial subjects were cued to perform one of two tasks on each trial, on half the trials the cue appeared before the target (750 SOA) and on the other half of trials the cue appeared at the same time as the target (0 SOA). In Experiment 1, nonword letter length and complexity did not interact with SOA when reading aloud. This suggests that target processing awaits the implementation of a task set. In Experiment 2 and 3, when only words were presented, word frequency and SOA interacted such that there was a smaller effect of word frequency at the 0 SOA relative to the 750 SOA. This suggests that lexical processing *can* go on in parallel with cue decoding. However, in Experiment 4, when words and nonwords were combined there was no interaction between word frequency and SOA. Participants appear to now wait to read the target word until the cue is processed, therefore reading aloud words *can* be affected by intention. Finally, in Experiment 5, when the task was to generate an antonym (instead of reading aloud), word frequency was additive with SOA. This suggests that at least some aspect of semantic processing (when generating an antonym) is delayed until cue processing is complete. These results, considered alongside results from the Psychological Refractory Period paradigm are taken to imply that sublexical processing uses both attention and intention. Critically, lexical and semantic processing may not require a form of attention, but they can be affected by intention. This runs contrary to the received view that reading aloud is automatic in the sense that it does not require intention, thus a radical change in how we think about the process of reading aloud is needed.

Acknowledgments

There are a number of people, without whom this thesis would not have been possible. First I must thank my advisor, Derek Besner, whose passion for this subject sparked my own while I was taking an undergraduate course with him many years ago. He has had an incredible impact on my career, and like a true mentor, on my life. His (occasionally abusively) constructive criticism has made me a better experimentalist, writer and teacher. I cannot express enough just how grateful I am to him. I also thank Michael Reynolds, who helped me in the transition from an undergraduate to a graduate student. He helped me chart the waters through my first research project and publication. To the 'senior grad' students in the Besner lab (who have all now graduated), I thank you for providing me with an intellectually stimulating environment, and for your friendships. Over the years I have had the pleasure of working with many research assistants and honors thesis students. I must thank them not only for the work they did (data collection and the like) but also for their curiosity. Helping someone discover and learn about psychology is one of the most enjoyable aspects of this work. Watching these students learn and grow as they begin their own paths in research has been incredibly rewarding.

When I was growing up, my parents impressed upon me the importance of finding a career that I loved. My mother, on countless occasions would advise my sisters and me to "find what you love to do, and figure out how to make it a career". I hope they know it is only because of their support and understanding that I have been able to find my passion. I must also thank my sisters, Amy and Christy, who have always supported me and I am proud to call my friends.

Finally, most importantly, I must thank my love, Tom. He has accepted my (somewhat unconventional) love of psychology and the geographic uncertainty that comes with this career without hesitation. For that, and for everything that he is, I will always love him.

Dedication

For Mom and Dad.

Table of Contents

List of Figures

List of Tables

Introduction

Understanding the basic processes involved in reading has been a goal of cognitive scientists since Huey (1908). The ability to turn symbols on a page into sound and meaning requires the coordination and execution of many cognitive sub-processes. My thesis focuses on a particular aspect of reading: the goal or intention of the reader (i.e., what they plan to do) and how this influences lexical and sublexical processing. A recurring theme that arises throughout my thesis is the effect of context. This will become clear in Experiment 4, which explores the hypothesis that the role of task set in reading aloud is not the same across all situations. I begin with a discussion of the dominant models of reading aloud and current theory regarding the role of "intention". I then outline the Task Set paradigm, which I use in all the experiments reported here.

Visual Word Recognition

Currently there are two general classes of visual word recognition models: localist dual-route models (e.g., Coltheart, Rastle, Perry, Langdon & Zeigler, 2001) and parallel distributed processing (PDP) dual-route models (e.g., Plaut, McClelland, Seidenberg & Patterson, 1996; Plaut, 2005). The present work is not designed to differentiate between these two classes of models. However, to simplify my thesis I have adopted the localist dual-route models as a framework in which to discuss my predictions and results, because they are currently more successful than any of the PDP models (e.g., Coltheart et al., 2001; Roberts, Rastle, Coltheart & Besner, 2003; Rastle, Havelka, Wydell, Besner & Coltheart, 2009). I return to the PDP models in the general discussion.

There are two major localist dual-route models of visual word recognition: the Dual Route Cascaded model (hereafter DRC; Coltheart et al., 2001) and the Connectionist Dual

Process model (hereafter CDP+; Perry, Ziegler & Zorzi, 2007). Both of these models have a dual route architecture consisting of a sublexical route, a lexical route and a semantic route (although the semantic route has not been implemented in either model to date; see Figure 1). Feature analysis across the letter string activates letter level analysis that in turn activates both sublexical and lexical routines.

The Sublexical Route

Following feature and letter activation, the sublexical route converts spelling to sound sub-lexically, by converting orthographic units into phonological units serially, from left to right. This route correctly reads aloud virtually all letter strings that could be words in terms of their orthography but happen not to be (e.g., “frane” and “frilp”) and all words that are regular in terms of their spelling-sound correspondences (e.g., gave/save/rave/wave and lint/mint/hint/dint). However, it assigns the regular pronunciation to strings like “have” and “pint”, rather than reading them aloud correctly.

There are some important differences in the details of exactly how the sublexical route functions in DRC and CDP+ and some of these differences are associated with differential success with respect to the ability to simulate some phenomena (e.g., consistency). The sublexical routine in CDP+ does make some peculiar errors that human do not make (as noted by Coltheart, 2011), and so will require some modifications to accurately capture human behaviour. However, these differences are not particularly germane to the present work, so I do not discuss them further.

The Lexical Route

The lexical route is identical in both models and consists of an Orthographic Input Lexicon (OIL) which contains a localist representation (lexical entry) for the spellings of each

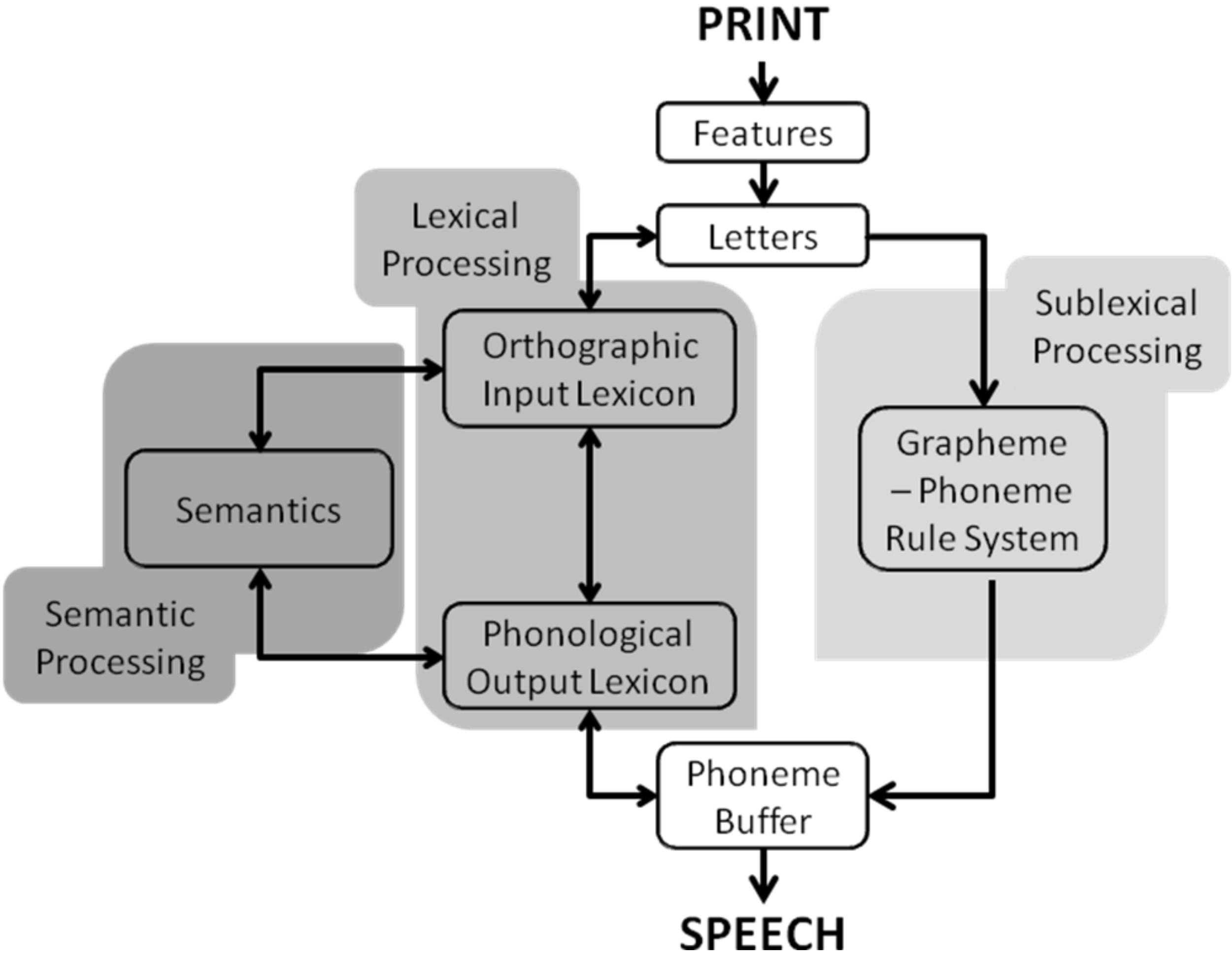


Figure 1. A schematic depiction of the DRC model of reading aloud, with lexical, sublexical, and semantic processing highlighted.

word a reader knows, and a Phonological Output Lexicon (POL) which contains a localist representation for the pronunciation of each word a person knows. Presentation of a word yields activation that cascades from features to letters through to the OIL and then to the POL and finally to the phoneme buffer. This route reads all words that have representations in both of these lexicons correctly, but is unable to read letter strings aloud that are not represented in these lexicons (e.g., “frane” and “frilp”).

The Semantic Route

Although these models are considered ‘dual’ route, there is a third route by which words can be read (though some might consider this a piece of the lexical processing route as opposed to a separate route). Semantic nodes receive activation from the OIL (at the same time activation is flowing forward, it also flows backward in an interactive fashion). The semantic level contains a representation for the meaning of each word a person knows. Activation is fed forward to the POL (again, activation also flows backward from the POL to semantics). To date, the semantic route has not been implemented in either DRC or CDP+. Regardless, semantics is a critical part of reading, as it is often the end goal (most people read for comprehension).

Automaticity

When a letter string is presented, activation flows along all routes in parallel. An assumption embedded in these models is that processing always unfolds in the same way; regardless of intention or context.[1] The researchers who developed these models do not explicitly state this, but there is little question that is how the models currently operate. This presumably

---

[1] This is also a true of PDP models (but see Kello & Plaut, 2003). It is of note that the CDP+ includes a role for spatial attention in the sublexical route (input occurs serially across a letter string, under the assumption that spatial attention is necessary for entry into the grapheme buffer).

developed out of the longstanding idea that reading aloud and semantic activation is "automatic". This notion can be attributed, in large part, to the results from the Stroop task, where subjects are presented a color word (e.g., "red" printed in BLUE) and asked to name the print color. When the print color is incongruent with the word, subjects are slower to respond than when the print colour is congruent with the word (see MacLeod, 1991 for a review). The fact that subjects appear to process the word, *even* when it hurts performance, has been widely accepted as evidence that they are incapable of preventing themselves from reading it. As Brown, Gore and Care (2002) noted "the Stroop task has provided influential data and the interpretation of that data has been that normal, mature word recognition is automatic" (p. 220). Thus, many researchers have claimed that reading is automatic.[2] Van Orden, Pennington and Stone (1990) stated that "through covariant learning, conscious rule application is replaced by the precise automatic phonologic coding that underlies skilled naming performance" (p. 510). Wilson Tregallas, Slason, Pasko & Rojas (2011) set out to "investigate the neurobiological substrates of the automatic, implicit stages of phonological processing" (p 724). Xu and Perfetti (1999) argued for "rapid automatic phonological activation, independent of stimulus based processing strategies" (p. 26). Brysbaert (2001) concluded that "phonological recoding in visual word recognition proves to be as mandatory in Dutch as in English" (p 772). Mari-Beffa, Valdea,

---

[2] There is a whole domain of research which argues that the reading deficits seen in individuals with developmental dyslexia or attention deficit hyperactivity disorder (ADHD) are due to these individuals failing to automatize the processes involved in reading, this is known as the General Automatization Deficit hypothesis (see Nicolson & Fawcett, 1990; Hurks et al., 2004) . This argument assumes that skilled readers process words automatically. For the most part researchers in this domain mean "automatic" in the sense that processing is fast and relatively easy; this is undoubtedly true for skilled readers. However, they also assume that reading is effortless and does not require attention in skilled readers. As such, the implications of any results that challenge this assumption will need to be considered, and a modification of this hypothesis may be called for.

Cullen, Catena & Houghton (2005) found results “supporting the view that this initial semantic access is automatic, occurring irrespective of what task is to be performed on the words” (p.301). Many other investigators make similar claims (Frost, 1998; Grainger, Diependaele, Spinelli, Ferrand, & Farioli, 2003; Johnston & Castles, 2003; Laberge & Samuels, 1974; Marcel, 1983; Neely, 1977; Neely & Kahan, 2001; Posner & Snyder, 1975; Reimer, Lorsbach & Bleakney, 2008; Xu & Perfetti, 1999, see Reynolds & Besner 2006 for more).

What exactly is meant by the term “automatic”? Automatic processes have been described as stimulus driven (Posner & Snyder, 1975), ballistic (i.e., once begun cannot be stopped; Besner, 2001; Hasher & Zacks, 1979) and unfolding independent of other processes (e.g., Brown, Gore & Carr, 2002; Logan, 1988). Along with these characteristics, almost all researchers agree that for a process to be automatic it does not require “attention” (Posner & Snyder, 1975; Logan 1988; Laberge & Samuels, 1974), central capacity (McCann, Remmington & Van Selst, 2000; Neely & Kahan, 2001), or intention (Posner & Snyder, 1975; Neely & Kahan, 2001; Brown, Gore & Carr, 2002). For example, Brown Gore and Carr described visual word recognition as

> “largely obligatory, in the sense that lexical processing is initiated by the presence of a word in the visual field, and largely autonomous, in the sense that once processing is initiated, lexical representations become sufficiently activated to influence decision and action. Processing to the lexical level tends to occur whenever words are present in a display, regardless of participants’ intention to read them, and often despite the intention to avoid reading them” (p 236).

Despite the plethora of researchers claiming that reading is automatic, there is a growing body of research that challenges this view. Melara and Mounts (1993) examined the discriminability of the word and color in the Stroop task (i.e., how easy it was to see/process the word vs. the color) and found that when discriminability was matched across the word and the color, the Stroop effect was largely reduced or eliminated with practice. They concluded that the Stroop effect is not mandatory; subjects do not read the word under all conditions (see also Sabri, Melara & Algom, 2001). Besner and Stolz (1999) demonstrated that spatial attention is necessary to obtain a Stroop effect. That is, if spatial attention is not distributed across the word, the Stroop effect is reduced. These results call into question the very source of the "reading is automatic" claim.

Outside of the Stroop task, it has been shown that reading aloud depends on spatial attention as a preliminary to lexical and semantic processing (Besner, Risko & Sklair, 2005; Risko, Stolz & Besner, 2011; Waechter, Besner & Stolz, 2011; Lachter, Forster & Ruthruff, 2004). As well, evidence from the Psychological Refractory Period (PRP) paradigm in which subjects are asked to perform two tasks in rapid succession, suggests that at least some processes underlying reading aloud require central capacity (Ruthruff, Allen, Lien & Grabbe, 2008; Cleland, Gaskell, Quinlan & Tamminen, 2006; Reynolds & Besner 2006; O'Malley, Reynolds, Stolz & Besner 2008; Besner, Reynolds & O'Malley 2009).

In short, research regarding whether or not reading aloud uses attention, intention and central capacity, under what conditions they use these resources and how processes involved in reading are affected by context, have only begun to be explored, but already a pattern is emerging: reading aloud is not automatic in the sense that it can be stopped and it does, for at least some processes, appears to require spatial and central attention.

There is some question, however, as to what constitutes evidence that processing is not automatic. Finkbeiner and Forster (2007) make the case that there are 3 general stages of processing: a pre-domain specific stage, a domain-specific stage and a decision (or response selection) stage. They argue that "to undermine the assumption that domain-specific processes proceed autonomously, one would need to demonstrate that the computations carried out at the second (domain-specific) stage of processing are directly modulated by higher-cognitive systems” (p. 59). Thus, spatial attention may be a necessary precursor to reading a word, but this likely affects a pre-domain specific stage (i.e., early perceptual processing) not the domain specific stage, and so does not constitute evidence that lexical and sublexical processing require attention. In contrast, evidence from the PRP paradigm, which shows that at least sublexical processing requires attention, suggests that the domain-specific processing of sublexical spelling-sound translation is not, contrary to Finkbeiner and Forster, “automatic” in the sense that it requires attention.

The aim of my thesis is to examine whether reading can be considered ‘automatic’ in the sense that it does not require *intention*. More specifically, I am interested in whether domain-specific stages of reading (sublexical, lexical and semantic processing) can be done in parallel with decoding a tone that tells subject which task to perform on each trial.

The Task Set paradigm

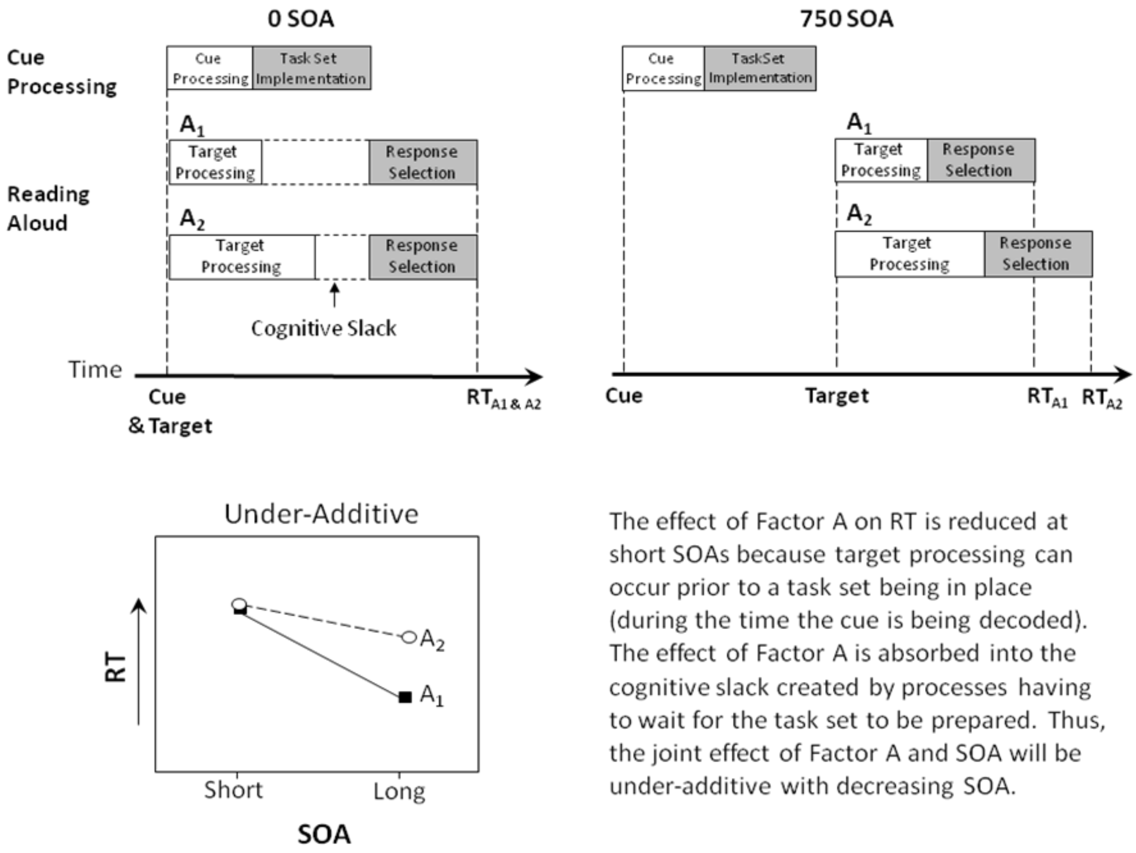
Besner and Care (2003) developed the Task Set paradigm.[3] On each trial, subjects are presented with a task cue indicating which of two tasks are to be performed. This task cue appears either at the same time as the onset of the target (zero SOA) or prior to it (i.e., -750 ms), and a factor that affects target processing time is also manipulated. When the task cue is presented 750 ms before the target there is enough time to process the cue and implement a task set prior to the onset of the target. In this case, the manipulated variable will yield a main effect (e.g., an effect of word

---

[3] In the original paper by Besner and Care (2003) and in those that followed (Besner & Risko, 2005; Ansari & Besner 2005) this paradigm was referred to as the “Task Choice” paradigm. Given that there is no ‘choice’ per se (subjects do not choose which task to perform) I now refer to this as the “Task Set” paradigm. Subjects are asked to switch between task sets throughout the experiment therefore this title better reflects the nature of the paradigm.

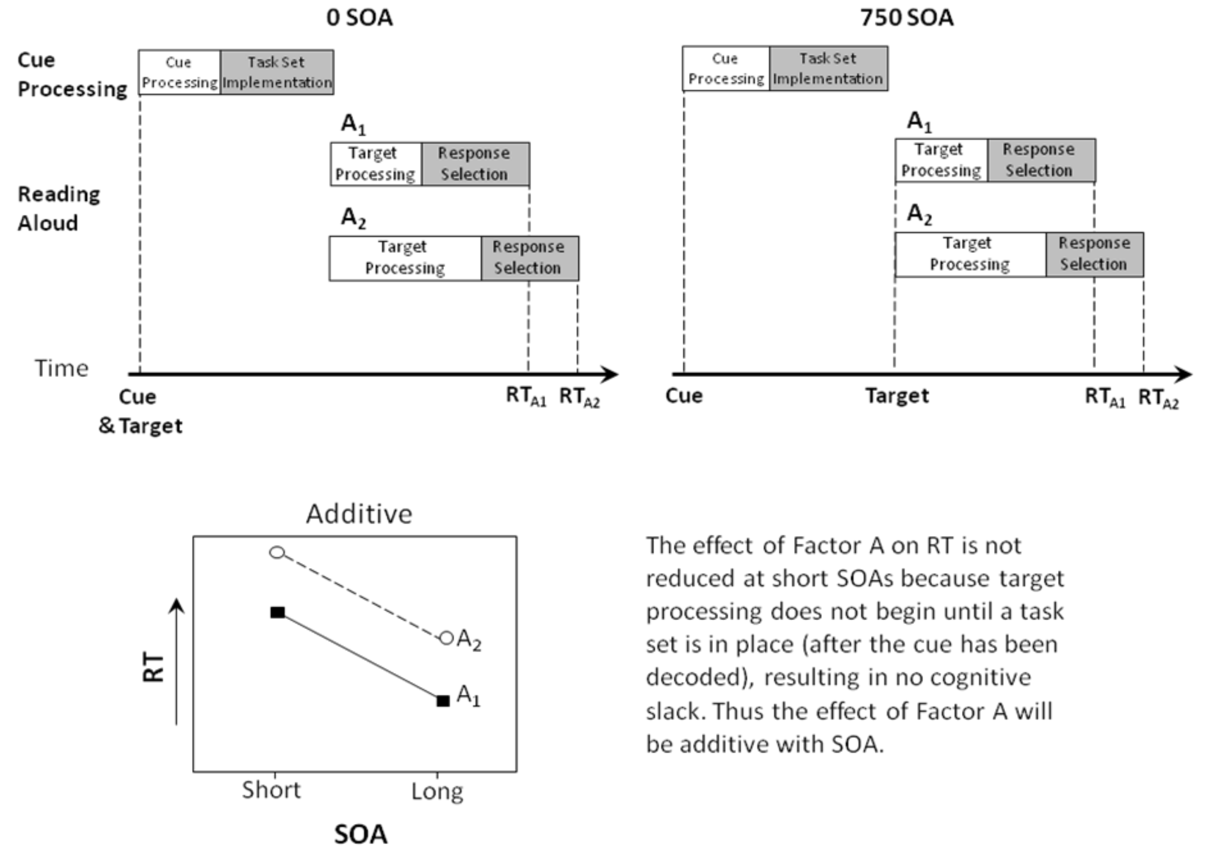
frequency). In contrast, when the SOA is zero, such that there is temporal overlap between the presentation of the task cue and the presentation of the target, the subject may either process the task cue and the target in parallel, or sequentially. If the subject can decode the task cue at the same time that the target is being processed then the effect of some manipulated target factor will not be observed, provided that the time to decode the cue is at least as long as the time taken to deal with the effect of that factor (see Figure 2). This is often described as processing of the task cue creating "cognitive slack" due to some later stage of processing being bottlenecked. This result (elimination of the effect of the manipulated factor) is expected if processing of the target is initiated without intention, is capacity free, ballistic and not interfered with by any other processes because the effect of that factor will be completely absorbed into the time taken to decode the task cue (hereafter, "absorbed into slack"). In contrast, if target processing does not begin until the task cue has been decoded (i.e., is bottlenecked until a task set is in place) then that factor and SOA should not interact because the stage at which the factor has its effect occurs after cue processing has been completed, and so the full effect of that factor will be observed (i.e., that factor and SOA will have additive effects on RT; see Figure 3).

Besner and Care (2003) argued that if one observed additivity between a manipulated variable and SOA, this should be taken as evidence that target processing (at least to the stage affected by the manipulated variable) required *intention*. That is, a task set must be in place before processing begins. However, this is not the only possible explanation for why processing of the target might be delayed. If cue processing (and task set implementation) and target processing both require some central resource, then processing of the target could be delayed until cue processing has finished (i.e., a limited capacity account). So long as cue processing, task set



The effect of Factor A on RT is reduced at short SOAs because target processing can occur prior to a task set being in place (during the time the cue is being decoded). The effect of Factor A is absorbed into the cognitive slack created by processes having to wait for the task set to be prepared. Thus, the joint effect of Factor A and SOA will be under-additive with decreasing SOA.

Figure 2. The standard account of an underadditive interaction between some manipulated factor (A) and SOA in the Task Set paradigm using cognitive slack logic. *RT_{A1}= Response time to Target A1; RT_{A2}= Response time to Target A2*



The effect of Factor A on RT is not reduced at short SOAs because target processing does not begin until a task set is in place (after the cue has been decoded), resulting in no cognitive slack. Thus the effect of Factor A will be additive with SOA.

Figure 3. The standard account of additive effects between some manipulated factor (A) and SOA in the Task Set paradigm using cognitive slack logic. *RT_{A1}= Response time to Target A1; RT_{A2}= Response time to Target A2*

implementation and target processing occur serially (regardless of the order) then additivity between SOA and the manipulated factor will be observed.[4] Thus, it could be that cue processing and target processing occur before task set implementation, but are done serially due to limited capacity.

If processing is delayed due to the need for some limited capacity resource, this would mean that processing requires attention, not intention. As noted by Besner and Care (2003) this account makes a unique prediction with regards to the second task. That is, if both tasks are being computed before the cue is processed a main effect of the manipulated factor should be seen across both tasks. I will address this issue in more detail in the General Discussion. For now, it is enough to say that I take additivity of a factor and SOA to indicate that cue processing and target processing do not unfold in parallel, and thus reading aloud is delayed due to the need for either attention or intention.

Besner and Care (2003) had subjects read nonwords aloud, and manipulated whether the target was clear or degraded (low stimulus quality). They used nonwords to determine whether sublexical processing could unfold in the absence of a task set because according to the two models of visual word recognition discussed above (DRC and CDP+) nonwords can only be read aloud correctly by recourse to sublexical processing (see Figure 1). Besner and Care found that stimulus quality and SOA did not interact. Following the logic outlined above, they took this result to imply that sublexical processing waits to begin until the cue has been decoded. Note,

---

[4] If cue processing and target processing both require a limited capacity resource and share resources such that they occur in parallel, but at a slower rate than when done independently then over-additivity will be observed (see Tombu and Jolicoeur, 2002 for a discussion of this issue in the context of the PRP paradigm). To date no one has observed this outcome in the Task Set paradigm and so I do not discuss it further.

however, that according to Finkbeiner and Forester (2007) this may simply mean that a pre-domain specific stage of processing is delayed, as the effect of stimulus quality may well be restricted to early perceptual processing. Support for this argument comes from the observation by Besner and Roberts (2003) that when reading nonwords aloud nonword letter length and stimulus quality had additive effects. According to additive factors logic (Sternberg 1969) this suggests that stimulus quality affects a different stage of processing than nonword letter length. Thus, it might be that the domain specific stage of sublexical processing is delayed until cue processing is completed only because a prior stage is delayed.

A related question is whether *lexical* processing when reading aloud can be carried out in the absence of a task set being in place. Paulitzki, Risko, O'Malley, Stolz and Besner (2009) and Kahan, Hengen & Mathis (2011) found that stimulus quality and SOA yielded an underadditive interaction, such that there was a smaller effect of stimulus quality at the 0 SOA as compared to the long SOA when *processing words*. This suggests that early processing of words (at least to the stage affected by stimulus quality) can go on in parallel with cue decoding. Paulitzki et al. (2009) also examined whether later lexical processing when reading aloud can unfold in the absence of a task set. They asked subjects to read aloud high and low frequency words *or* make a case decision (upper vs. lower case) decision to the words. They reported no significant interaction between word frequency and SOA. However, they did see a small trend (9 ms) towards underadditvity. Paulitzki et al. argue that the observed additivity between word frequency and SOA indicates that the implementation of a task set is a preliminary to lexical processing when reading aloud.

Experiment 1 investigates, more directly than in Besner and Care, whether sublexical processing can be carried out in parallel with cue decoding. In Experiments 2 and 3 I re-examine whether lexical processing can go on in parallel with cue processing, and find very different results from those reported by Paulitzki et al. (2009). In Experiment 4 words and nonwords are combined so as to determine if the nature of the stimulus set has any impact on whether a task set must be in place before processing begins. Finally, Experiment 5 examines whether a particular form of semantic processing (that involved in generating an antonym) can occur prior to a task set being in place. Overall the results reported here support the conclusion that reading aloud is not “automatic” in the sense that the domain specific stages involved are affected by intention or require some form of attention.

Experiment 1: Sublexical Processing

Experiment 1 investigates whether sublexical processing can be carried out in parallel with cue decoding. To date, the only factor that has been used to examine sublexical processing in the Task Set paradigm is stimulus quality, which arguably, has much of its effects early in processing (Besner & Care, 2003; Besner & Roberts, 2003; Kahan et al, 2010; but see Yap & Balota, 2007; O'Malley & Besner, 2008). To investigate this issue more directly I manipulated the letter length of nonwords because (i) it is well known that as the number of letters increases, the longer it takes to begin to pronounce it and (ii) this effect has been attributed to sublexical processing (see Weekes, 1997; Coltheart et al, 2001; Perry et al, 2007; Roberts, Rastle, Coltheart & Besner, 2003). Following Reynolds and Besner (2006), nonword complexity was deliberately confounded with letter length so as to increase the effect size of the sublexical manipulation. Here, complexity refers to the presence/absence of "whammies" in a nonword. That is, a whammy is present when a letter modifies the pronunciation of the previous letter (e.g., 'oo' and 'ph' in "fooph"); RTs are slower to "fooph" as compared to when whammies are not present (e.g., as in "frulp"; see Rastle & Coltheart, 1998). With letter length and complexity confounded, the main effect of this factor is larger, which in turn increases the power to detect an interaction with SOA.

Method

Subjects. Forty-eight undergraduate students were recruited from the Psychology undergraduate student subject pool at the University of Waterloo. Each subject was awarded credit towards one of their courses for their participation in a single session lasting 25-30 minutes. All subjects reported English as their first language and had normal or corrected-to-normal vision.

Stimuli. The stimulus set consisted of 104 short simple nonwords and 104 long, complex nonwords that were used by Reynolds and Besner (2006) in their Experiments 3 and 4. The stimuli appear in the Appendix A. An additional 16 nonwords with similar characteristics (8 short, simple nonwords and 8 long, complex nonwords) were used during the practice trials. The stimulus set for the experiment proper was divided into four: half the items were assigned to the reading aloud task and the half to the case decision task. Half of the items within each task were assigned to the 0 SOA condition and the other half where assigned to the 750 SOA condition. This assignment of items to tasks and SOA was counterbalanced across subjects; subjects were assigned to a stimulus list counterbalance based on order of arrival in the lab.

Task cues. On each trial a 100 ms tone indicated which of the two tasks was to be performed. For half the subjects a high frequency tone (2500 Hz) required subjects to read the word aloud, while a low tone (500 Hz) required subjects to respond to the case the word was presented in; these were the same tones used by Paulitzki et al. (2009). The remaining subjects were assigned to the reversed tone-task mapping. Subjects were assigned to a cue counterbalance based on order of arrival in the lab.

Apparatus. The experiment was conducted using E-Prime software running on a Pentium-IV computer. E-Prime software controlled the timing and presentation of stimuli and logged response times. Stimuli were presented to subjects on a standard 17”-inch SVGA color monitor. Subject’s vocal responses were collected using an Altec-Lansing microphone.

Design. The design was a 2 x 2 x 2 factorial in which the factors were Task (Reading Aloud vs. Case Decision), SOA (0 vs. 750 ms) and nonword complexity (short, simple vs. long, complex). All conditions were randomly intermixed within a single block of trials. Each subject received a different random sequence.

Procedure. Subjects were tested individually in a dimly lit room and were seated in front of the computer monitor and given written and verbal instructions. Subjects were told that on each trial a tone would sound and a string of letters would appear in the centre of the computer screen. The instructions stated that, depending on the tone (as outlined above), the task was to either pronounce the string of letters presented on the screen (Reading Aloud Task) or to indicate what case the item was presented in by saying “upper” or “lower” (Case Decision Task). Responses activated a microphone connected to a voice-key assembly that recorded reaction times.

Each trial began with the presentation of a fixation cross for 500 ms, followed by a 100 ms duration tone. In the advance cue condition the target word appeared 750 ms after the onset of the tone. In the zero SOA condition the target word appeared at the same time as the onset of the tone. Following the subject’s response, a blank screen was displayed until the experimenter logged the subject’s response as correct or not. Once the experimenter keyed in a response, an inter-trial interval of 500 ms ensued. Responses were classified in four ways: (1) a spoiled trial was logged if the microphone did not pick up the subject’s vocal response, or some external noise triggered the microphone before the subject could respond; (2) incorrect (mispronounced), (3) incorrect task and (4) a correct response.

All subjects performed one practice block of 16 trials prior to the start of the experimental trials. During practice trials, subjects were given feedback on their performance. If the subject made an error, they were told specifically what type of error (e.g. mispronounced the target or performed the wrong task). No feedback was given during the experiment proper.

Results

Linear mixed effects modeling allows for both subjects and items to be used as crossed-random effects; separate subject and item analysis are therefore not necessary. To supplement this analysis, I also provide the vincentile plots, which show the RT distribution for the joint effects of word frequency and SOA (e.g., see Yap, Balota, Tse & Besner, 2008).

I report $\beta^{\wedge}$, the corresponding *t* value for RTs, and the *z* value for errors (one advantages of this method is that t-values and z-values can be interpreted in the standard way). Markov chain Monte Carlo sampling of the posterior distribution of the parameters was used in the RT analysis to analyze significance (p-values). I complete the specification by reporting the standard deviation of the random effects. All analyses were computed in R, using the lme4 package (R Development Core Team, 2004).

Mean RTs and mean percentage errors for each condition of Experiment 1 can be seen in Table 1. RTs and errors for the words were fitted to linear mixed-effects models, as outlined by Baayen (2009), with subject and items as crossed random effects. Responses to each task were fitted to different models (i.e., one model was tested for the reading aloud task and a separate one was tested for the case decision task). The following factors were included in all initial models: SOA, Nonword Complexity, Previous RT (RT on trial N-1), Task switch/repetition (whether the task to be performed on the current trial was the same as the previous trial or different), counterbalance and trial number. Including the control predictors of previous RT, counterbalance and trial number allows the model to account for some of the variance that would otherwise be considered error variance. Using the most elaborate model also reduces the potential for bias in the results (one could test several different models, and select the one that suits them best – using the "include all variables" rule-of-thumb prevents this).

	Reading Aloud				Case Decision			
	RTs		% Error		Case Decision		% Error	
	0 SOA	750 SOA	0 SOA	750 SOA	0 SOA	750 SOA	0 SOA	750 SOA
Complex	1026	744	5.8	3.6	1039	679	6.7	4.2
Easy	985	683	3.1	3.3	1026	686	6.7	4.0
Difference	**41**	**61**	**2.7**	**0.3**	**13**	**-7**	**0.0**	**0.2**

Table 1. Mean Reaction Times (ms) and Percentage Errors (% Errors) in Experiment 1 as a function of Task, SOA, and Nonword Complexity.

Incorrect responses (4.7%) and spoiled trials (6.2%) were discarded prior to the RT analysis. The RTs were first fitted to a linear mixed effect model that contained only the main effects of each factor. RTs that were greater than 2.5 standard deviations from the predicted RTs of the model were removed prior to further analysis resulting in an additional 2.3% of the correct RTs being removed from the Reading Aloud task and 2.1% of the correct RTs being removed from the Case Decision task.

Reading Aloud RTs. There was a main effect of nonword complexity, $\beta^{\wedge}$ = 66.3, *t(4376)* = 5.4, *p* < .01, and a main effect of SOA, $\beta^{\wedge}$ = 257.8, *t(4376)* = 18.3, *p* < .01. There was also a main effect of previous RT, $\beta^{\wedge}$ = *0.06*, *t(4376)* = 11.1, *p* < .01, a main effect of trial, $\beta^{\wedge}$ = *-0.5*, *t(4376)* = -8.1, *p* < .01, and of counterbalance, $\beta^{\wedge}$ = *-90.9*, *t(4376)* = -8.9, *p* < .01. The main effect of task switch/repetition was not significant, $\beta^{\wedge}$ = *17.4*, *t(4376)* = 1.6, *p* > .05 but did interact with SOA such that the effect of SOA was greater on task switch trials than on task repetition trials $\beta^{\wedge}$ = *62.4*, *t(4376)* = 14.9, *p* < .01. Critically, the interaction between nonword complexity and SOA was not significant, $\beta^{\wedge}$ = *18.2*, *t(4376)* = 1.2, *p* > .05; the effect of complexity was the same at the 0 SOA and at the 750 SOA. Note that although this interaction is not significant, there is an underadditive trend of 20 ms in the means; this issue is addressed further when considering the error data. The standard deviation of the random effect of item was estimated at 45.9. The standard deviation of the by-subject adjustments was estimated at 269.5. The residual standard deviation was 242.1.

Reading Aloud Errors. There was no main effect of nonword complexity, $\beta^{\wedge}$ = .16, *z* = .6, *p* > .05, no main effect of SOA, $\beta^{\wedge}$ = .38, *z* = 1.2, *p* > .05 nor of task switch/repetition, $\beta^{\wedge}$ = -.34, *z* = -1.43, *p* > .05. There was a significant main effect of previous RT, $\beta^{\wedge}$ = -.0003, *z* = -3.6, *p* < .01 and of trial, $\beta^{\wedge}$ = .003, *z* = 2.5, *p* < .05. Interestingly there was a significant interaction between SOA and nonword complexity, $\beta^{\wedge} = -.67$, $z = -2.1$, $p < .05$, such that there was a *larger* effect of complexity at the 0 SOA than at the 750 SOA. This over-additive interaction suggests there is a speed-accuracy trade-off in which subjects are responding faster to the complex items at the 0 SOA, but are making more errors; this provides an explanation for the (non-significant) underadditive trend seen in the mean RTs. The standard deviation of the random effect of item was estimated at 1.1. The standard deviation of the by-subject adjustments was estimated at 1.4.

Case Decision RTs. There was no main effect of nonword complexity, $\beta^{\wedge} = .4$, $t\,(4619) = -.04$, $p > .05$. There was a main effect of SOA, $\beta^{\wedge} = 296.2$, $t(4619) = 25.9$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = 0.09$, $t(4619) = 12.0$, $p < .01$, and of trial, $\beta^{\wedge} = -0.5$, $t(4619) = -7.4$, $p < .01$. There was a main effect of task switch/repetition, $\beta^{\wedge} = 73.5$, $t(4619) = 6.4$, $p < .01$, which interacted with SOA, $\beta^{\wedge} = 90.3$, $t(4619) = 5.6$, $p < .01$ such that the effect of SOA was greater on task switch trials than on task repetition trials. There was no interaction between SOA and nonword complexity. The standard deviation of the random effect of item was estimated at 27.9. The standard deviation of the by-subject adjustments was estimated at 185.3. The residual standard deviation was 269.8.

Case Decision Errors. There was a main effect of SOA, $\beta^{\wedge} = .7$, $z = 2.9$, $p < .05$, a main effect of task switch/repetition, $\beta^{\wedge} = -.65$, $z = -2.9$, $p < .05$, and a main effect of trial, $\beta^{\wedge} = .006$, $z = 5.4$, $p < .01$. None of the other main effects (Complexity, previous RT and counterbalance) or interactions approached significance. The standard deviation of the random effect of item was estimated at .2. The standard deviation of the by-subject adjustments was estimated at .8.

Vincentiles. A vincentizing procedure was used to generate a response time distribution (Vincent, 1912). Ten vincentiles (the mean of observations within a given percentile range) were first computed for each subject. The individual vincentiles were then averaged across subjects and the mean vincentiles plotted. The mean vincentiles are plotted as a function of nonword complexity and SOA for the reading aloud condition and can be seen in Figure 4. As there was no main effect of nonword complexity in the case decision task (and it is not the task of interest), I do not report the vincentiles for the case decision task here or for Experiments 2 through 4.

The vincentile distribution can be seen in Figure 4. At the 750 SOA the effect of nonword complexity increases throughout the distribution, as expected. At the 0 SOA, the distribution has a curved appearance, over the fastest five vincentiles the effect of nonword complexity is increasing, but in the slowest five vincentiles it is decreasing. This pattern can be explained by the speed accuracy trade-off, assuming that this trade-off affects the slowest trials more. That is, if there was no speed accuracy trade-off, the effect of nonword complexity would likely increase throughout the distribution (much like in the 750 SOA condition). This explanation will need to be supported by a replication in which there is no speed accuracy trade-off (which can be seen in Experiment 4).

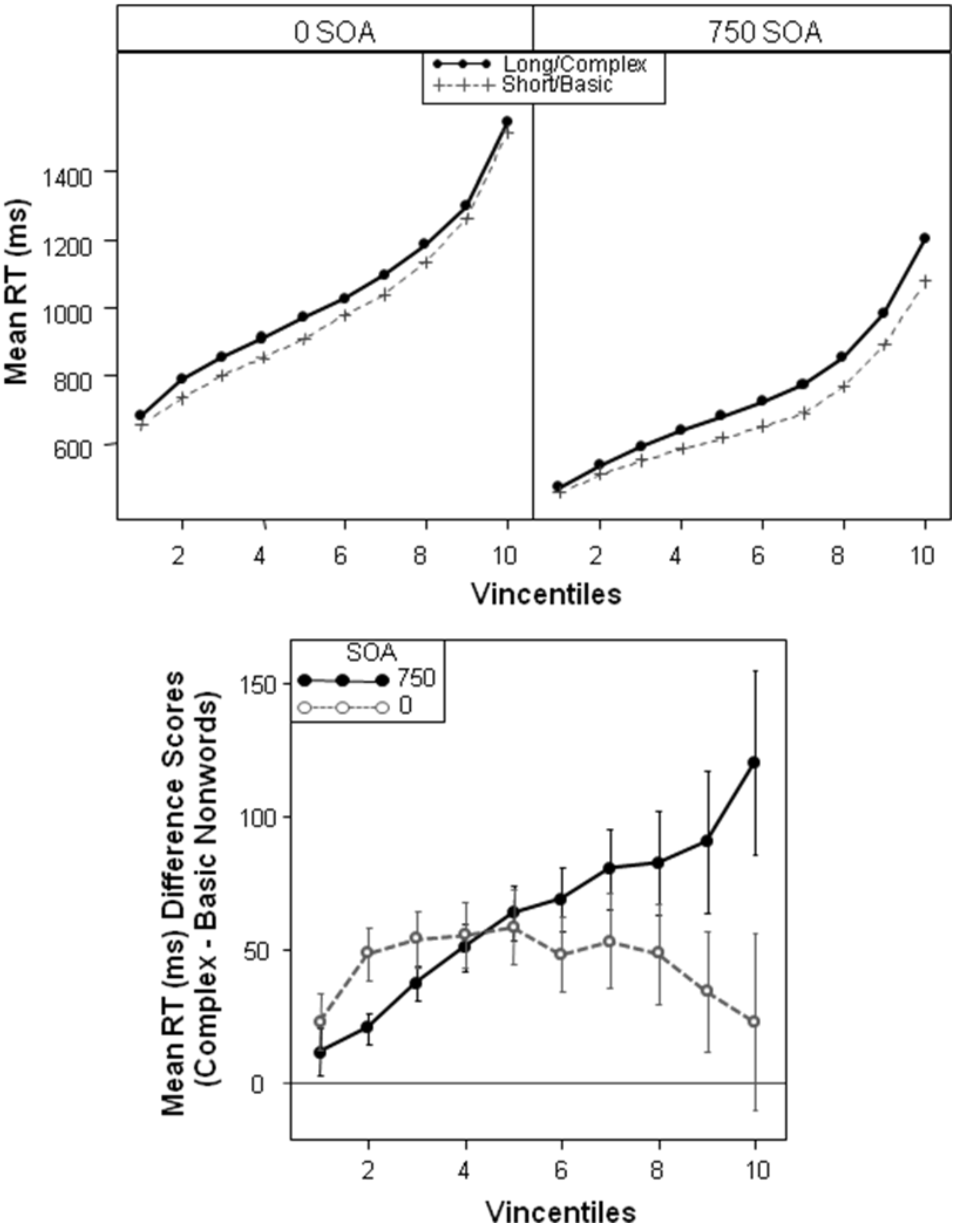


Figure 4. Top panel: Vincentile means for participant's reading aloud RTs in Experiment 1 as a function of SOA and nonword complexity. Bottom panel: The difference in vincentile means for basic versus complex nonwords for participants reading aloud RTs. Vertical bars represent the standard error of the mean.

Discussion

The results of Experiment 1 are clear: there was no significant under-additive interaction between nonword complexity and SOA in either RTs or errors. The slight (non-significant) trend towards underadditivity in the mean RTs can be explained by the significant overadditive interaction in the errors (suggesting a speed accuracy trade-off). This result is entirely consistent with the claim that sublexical processing does not begin until after the task cue has been processed (related, see Besner & Care, 2003; Kahan et al., 2010). However, the empirical basis for this claim is stronger in the present data than in Besner and Care's experiment given the manipulation of a factor that directly affects sublexical processing. This result provides support for the argument that sublexical processing requires either attention or intention. Although nonword letter length is thought to directly index the domain specific stage of sublexical spelling-to-sound conversion, it is possible that earlier processing (e.g., feature and letter processing) is the stage that cannot go on in parallel with cue processing (sublexical processing is only delayed because this earlier processing is delayed). However, if early feature and letter level processing is not delayed by cue processing, then it is likely that the domain-specific stage of sublexical processing is delayed until cue processing is complete. To preview the results of Experiments 2 and 3, we see an underadditive interaction between word frequency and SOA when reading words. Given that feature and letter analysis is common to both lexical and sublexical routines, this provides some evidence that feature and letter level processing can go on in parallel with cue processing. Thus, sublexical processing (but not earlier stages of processing) does not begin until the cue has been processed. I will return to the issue of whether *intention* or *attention* is the critical factor in the General Discussion.

An interaction between SOA and task switching was observed, such that there was a larger effect of task switching at the 0 SOA than at the 750 SOA. This can be explained by priming associated with decoding the task cue on task repetition trials (here task cue and task repetition are confounded) such that there is a greater benefit to repeating the task cue when the SOA is short than when the SOA is long. This is hardly surprising; more priming is expected at the zero SOA than the long SOA, as priming surely decays with time. Note there is no three-way interaction between SOA, task switching and nonword complexity, suggesting that the two way interaction between SOA and switch/repeat does not reflect target processing. The argument applies to the observation of this interaction in all the subsequent experiments.

Experiment 2: Lexical Processing

Paulitzki et al. (2009) examined whether lexical processing when reading aloud can unfold in the absence of a task set by having subjects either read aloud high and low frequency words *or* make a case decision (upper vs. lower case). At the long SOA (750 ms) subjects have ample time to decode the task cue before being presented with the target. Consequently, high frequency words were read aloud faster than the low frequency words. Most investigators would also expect that there would be no effect of word frequency in the case decision task given that it does not require lexical level processing (the level at which word frequency is typically considered to have its effect). Paulitzki et al. found no interaction between SOA and word frequency (the effect of word frequency was statistically the same size at the 0 SOA as at the 750 SOA). However, there was a non-significant 9 ms trend towards underadditivity. They took these data to imply that *lexical* processing does not unfold in parallel with cue decoding.

It is important to note that Paulitzki et al.'s conclusions rely on accepting the null hypothesis of no interaction between SOA and word frequency. If it is generally true that word frequency and SOA are additive factors in the context of the task set paradigm then a radical change in current theory is called for. However, it is not clear what specific modifications are needed. For example, should we adopt the assumption that a task set must be in place before target processing can begin, or is it that there is interference between cue decoding and target processing (or both). Before attempting to determine which particular assumption is problematic it is prudent to determine just how easy it is to replicate Paulitzki and colleagues' finding. In Experiment 2 the tasks were the same as in Paulitzki et al. but the response mode in the case decision task was vocal rather than manual. It is possible that having both tasks use the same response modality (here, vocal) would allow for concurrent processing of the task cue and the target. This possibility is plausible given that response mode affects the size of the task switching effect (when the response mode for two tasks are similar across tasks, there is a smaller switch cost than when the response modes are different, Arrington, Altmann & Carr, 2003). In Experiment 3 the tasks and responses were the same as in Paulitzki et al. The only other modification was the use of a different word set in both experiments reported here. Specifically, a word set was selected that had a larger range of word frequencies, included monosyllabic and polysyllabic words, and ranged from 3 to 9 letters in length. This word set better represents the range of words people typically encounter in text than does the 4- to 5 letter monosyllabic words used by Paulitzki and colleagues (though this does not imply that the current set of words is easier to read). To anticipate the central result, in direct contrast to the results observed by Paulitzki et al., both experiments yielded a strong underadditive interaction between word frequency and SOA. That is, the word frequency effect was significantly smaller at the 0 SOA than at the 750 SOA.

Method

Subjects. Forty-two undergraduate students participated in Experiment 2. All subjects were recruited from the Psychology undergraduate student subject pool at the University of Waterloo. Each subject was awarded credit towards one of their courses for their participation in a single session lasting 25-30 minutes. All subjects reported English as their first language and had normal or corrected-to-normal vision.

Stimuli. The stimulus set was selected from words used in the English Lexical Project (Balota et al. 2007) and consisted of 144 high frequency words (Log HAL frequency norms[5] between 4.6 and 9.7, with an average of 7.9) that varied in length from 3 to 8 letters (average

---

[5] Log HAL frequency norms are the log transformed norms based on the Hyperspace Analogue to Language (HAL) corpus (Lund & Burgess, 1996).

length of 5.8) and 144 low frequency words (Log HAL frequency between 9.7 and 14.3, with an average of 11.2) that varied in length from 3 to 9 letters (average length 5.6). The stimuli appear in the Appendices B and C. An additional 16 words with similar characteristics (8 high frequency, and 8 low frequency) were used during practice.

The words used here were selected for use in Experiments 2 -5 (i.e., all of the experiments that used word stimuli). To be useable in the antonym generation task of Experiment 5, I only selected words that had an obvious antonym. The strength of association between the target and its most strongly associated antonym was approximately matched across high frequency and low frequency items using the Nelson, McEvoy and Schreiber (2004) strength of association norms (.3 for high frequency words and .2 for low frequency words). The antonyms used to calculate the association norms can also be seen in Appendices B and C.

The stimulus set for the experiment proper was divided into four: half the items were assigned to the reading aloud task and the half to the case decision task. Half of the items within each task were assigned to the 0 SOA condition and the other half where assigned to the 750 SOA condition. This assignment of items to tasks and SOA was counterbalanced across subjects; subjects were assigned to a stimulus list counterbalance based on order of arrival in the lab.

Design. The design consisted of a 2 x 2 x 2 factorial in which the first factor was Task (Reading Aloud vs. Case Decision), the second factor was SOA (0 vs. 750 ms) and the third factor was word frequency (High vs. Low). All conditions were randomly intermixed within a single block of trials. Each subject received a different random sequence.

Procedure. The procedure, apparatus and task cues were identical to that of Experiment 1, except subjects were told that they would always be presented with a word.

Results

Mean RTs and percentage errors for each condition for Experiment 2 can be seen in Table 2. RTs and errors for the words were fitted to linear mixed-effects models with subject and items as crossed random effects. Responses to each task were fitted to different models. The following factors were included in all initial models: SOA, word frequency, Previous RT (RT on trial N-1), Task switch/repetition (whether the task to be performed on the current trial was the same as the previous trial or different), counterbalance and trial number. The factors of counterbalance and trial number were not significant in any of the models and so were removed from the final models.

Incorrect responses (1.9%) and spoiled trials (2.7%) were discarded prior to the RT analysis. The RTs were first fitted to a linear mixed effect model that contained only the main effects of each factor. RTs that were greater than 2.5 standard deviations from the predicted RTs of the model were removed prior to further analysis resulting in an additional 2.4% of the correct RTs being removed from the Reading Aloud task and 2.2% of the correct RTs being removed from the Case Decision task.

Reading Aloud RTs. There was a main effect of word frequency, $\beta^{\wedge} = 47.1$, $t(5604) = 4.9$, $p < .01$, and a main effect of SOA, $\beta^{\wedge} = 306.3$, $t(5604) = 14.4$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = 0.08$, $t(5604) = 12.4$, $p < .01$, and of task switch/repetition, $\beta^{\wedge} = 30.6$, $t(5604) = 3.9$, $p < .05$. Critically, the underadditive interaction between word frequency and SOA was significant $\beta^{\wedge} = -33.8$, $t(5604) = -2.6$, $p < .05$, such that there was a smaller effect of word frequency at the 0 SOA than at the 750 SOA. There was also a significant interaction between SOA and task switch/repetition, $\beta^{\wedge} = 101.7$, $t(5604) = 9.2$, $p < .01$, in which the effect of SOA was smaller when the previous task was the same as the current task, relative to the size of the

	Reading Aloud				Case Decision			
	RTs		% Error		RTs		% Error	
	0 SOA	750 SOA	0 SOA	750 SOA	0 SOA	750 SOA	0 SOA	750 SOA
Low frequency	968	675	2.6	3.1	1067	754	2.1	1.6
High Frequency	960	629	1.1	1.1	1055	750	2.2	1.1
Difference	**8**	**46**	**1.5**	**2.0**	**12**	**4**	**-0.1**	**0.5**

Table 2. Mean Reaction Times (ms) and Percentage Errors (% Errors) in Experiment 2 as a function of Task, SOA, and Word Frequency.

SOA effect when the previous task was different. There was no significant interaction between task switch/repetition and word frequency. The standard deviation of the random effect of item was estimated at 36.1. The standard deviation of the by-subject adjustments was estimated at 152.9. The residual standard deviation was 204.7.

Reading Aloud Errors. There was a main effect of word frequency, $\beta^{\wedge}$ = -1.2, *z* = -2.0, *p* = .05 and a main effect of previous RT, $\beta^{\wedge}$ = -.0005, *z* = -2.5, *p* < .05. No other main effects or interactions approached significance. The standard deviation of the random effect of item was estimated at 3.4. The standard deviation of the by-subject adjustments was estimated at 2.4.

Case Decision RTs. There was no main effect of word frequency, $\beta^{\wedge}$ = 10.3, *t (5668)* = 1.6, *p* > .05. There was a main effect of SOA, $\beta^{\wedge}$ = 237.4, *t (5668)* = 11.0, *p* < .01. There was also a main effect of previous RT, $\beta^{\wedge}$ = *0.05*, *t(5668)* = 3.3, *p* < .05, and of task switch/repetition, $\beta^{\wedge}$ = *38.4*, *t(5668)* = 2.0, *p* = .05. The interaction between SOA and task switch/repetition was significant, $\beta^{\wedge}$ = *57.2*, *t(5668)* = 2.1, *p* < .05. There was also a significant interaction between SOA and previous RT, $\beta^{\wedge}$ = *0.05*, *t(5668)* = 2.3, *p* < .05. Finally, there was a significant interaction between task switch/repetition and previous RT, $\beta^{\wedge}$ = *0.05*, *t(5668)* = 2.3, *p* < .05, however the three way interaction between SOA, task switch/repetition and previous RT was not significant $\beta^{\wedge}$ = *-0.01*, *t(5668)* = -.04, *p* > .05. The standard deviation of the random effect of item was estimated at 19.3. The standard deviation of the by-subject adjustments was estimated at 177.3. The residual standard deviation was 231.1.

Case Decision Errors. None of the main effects (SOA, frequency, Task switch/repetition or previous RT) were significant. There was an interaction between task switch/repetition and previous RT, $\beta^{\wedge}$ = -.002, *z* = -2.2, *p* < .05 and a marginal interaction between task switch/repetition and SOA, $\beta^{\wedge}$ = -1.9, *z* = -1.7, *p* = .07. No other interactions approached significance. The standard deviation of the random effect of item was estimated at 3.5. The standard deviation of the by-subject adjustments was estimated at 3.2.

Vincentiles. If SOA and word frequency are generally under-additive (i.e., not just in the means) then the effect of word frequency should be smaller in the 0 SOA condition relative to the 750 SOA condition throughout the distribution. The difference scores (low frequency – high frequency; see bottom panel of Figure 5) for the reading aloud condition show that this is the case, and that the effect of word frequency increases with increasing vincentiles in the 750 SOA condition, but not in the 0 SOA condition. It should be noted, however, that the effect of word frequency does not center on zero in the 0 SOA condition, suggesting that there is some residual effect of word frequency.

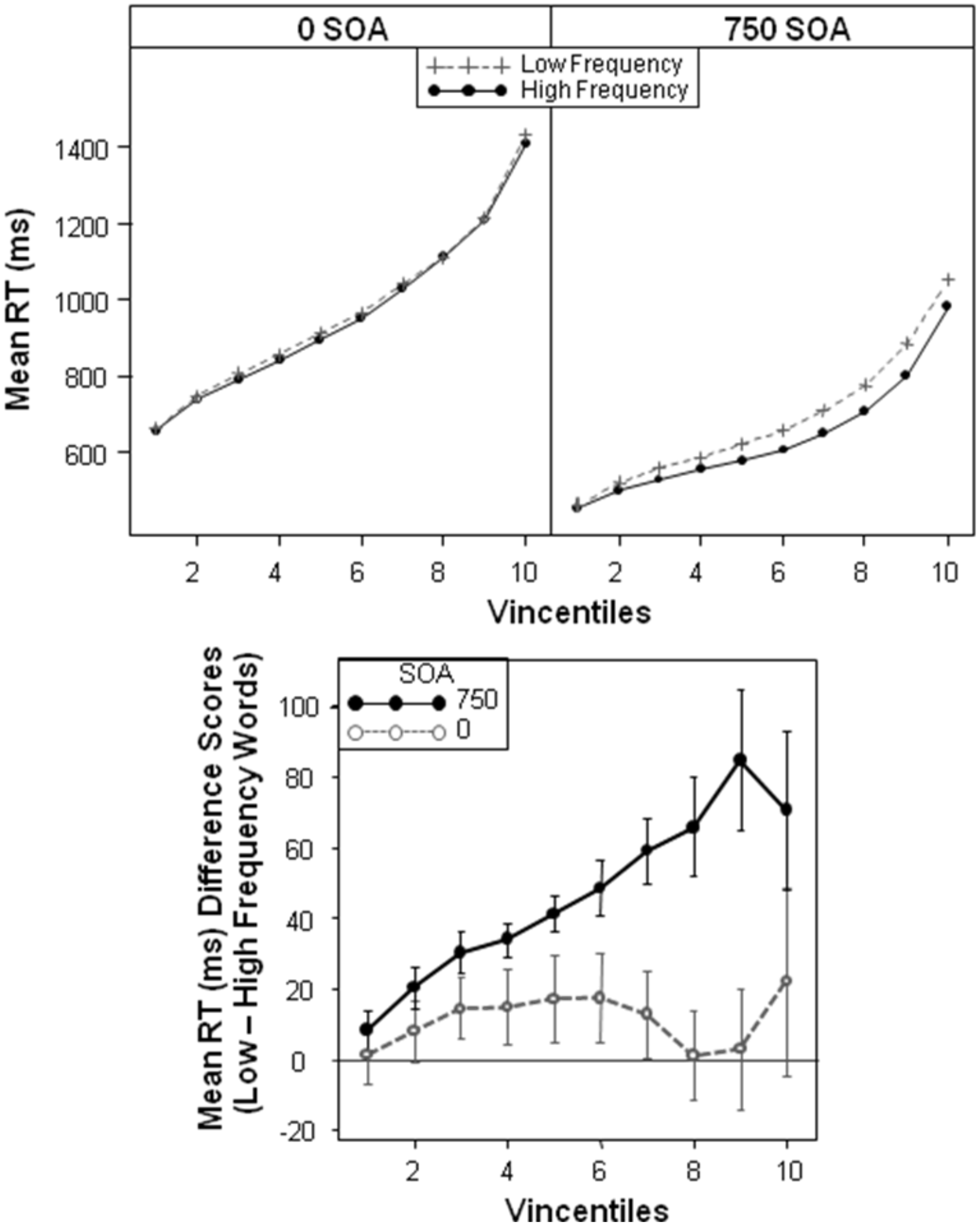


Figure 5. Top panel: Vincentile means for participant’s reading aloud RTs in Experiment 2 as a function of SOA and word frequency. Bottom panel: The difference in vincentile means for low versus high frequency items for participant’s reading aloud RTs. Vertical bars represent the standard error of the mean.

Experiment 3: Lexical Processing

An interaction between SOA and word frequency was observed in Experiment 2 unlike in the experiment reported by Paulitzki et al. (2009), who observed no significant interaction between these factors. The present result is more likely to reflect the true state of affairs than the null result reported by Paulitzki et al. (a Type 1 error is less likely than a Type 2 error). Nonetheless, it is important to ensure that the difference in results was not simply due to the difference in response mode for the case decision task (in Paulitzki et al. the response was made via a key press, but here it was made vocally). Therefore, Experiment 3 is identical to Experiment 2 with the exception that responses to the case decision task were made by key press.

Method

Subjects. Sixty subjects, recruited from the Psychology undergraduate student subject pool at the University of Waterloo, participated. Each subject was awarded credit towards one of their courses for their participation in a single session lasting 25-30 minutes. All subjects reported English as their first language and had normal or corrected-to-normal vision.

Procedure. The procedure, apparatus, stimuli and design were identical to those of Experiment 2, with one exception: the case decision task. Here subjects pressed the ‘G’ button on the keyboard if the word was in uppercase and ‘H’ if it was in lower case.

Results

Mean RTs and errors for Experiment 3 can be seen in Table 3. RTs and errors for the words were fitted to linear mixed-effects models with subject and items as crossed random effects. Responses to each task were fitted to different models. The following factors were included in all initial models: SOA, word frequency, Previous RT (RT on trial N-1), Task switch/

	Reading Aloud				Case Decision			
	RTs		% Error		RTs		% Error	
	0 SOA	**750 SOA**	**0 SOA**	**750 SOA**	**0 SOA**	**750 SOA**	**0 SOA**	**750 SOA**
Low frequency	832	580	3.7	2.2	894	582	5.6	2.7
High Frequency	816	545	3.3	1.4	900	578	5.6	3.1
Difference	**16**	**35**	**0.4**	**0.8**	**-6**	**4**	**0.0**	**-0.4**

Table 3. Mean Reaction Times (ms) and Percentage Errors (% Errors) in Experiment 3 as a function of Task, SOA, and Word Frequency.

repetition (whether the task to be performed on the current trial was the same as the previous trial or different), counterbalance and trial number. The factors of counterbalance and trial were not significant in any of the models and so were dropped from the final models.

Incorrect responses (3.5%) and spoiled trials (2.7%) were discarded prior to the RT analysis. Outliers were removed using the same trimming procedure as in Experiment 1 resulting in an additional 2.1% of the correct RTs being removed in the reading aloud task and 2.1% being removed from the Case Decision task.

Reading Aloud RTs. There was a main effect of word frequency, $\beta^{\wedge} = 37.4$, $t(7969) = 5.2$, $p < .02$, and a main effect of SOA, $\beta^{\wedge} = 248.0$, $t(7969) = 19.5$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = 0.07$, $t(7969) = 15.5$, $p < .01$, and of task switch/repetition, $\beta^{\wedge} = 19.5$, $t(7969) = 3.5$, $p < .05$. Critically, the under-additive interaction between word frequency and SOA was significant $\beta^{\wedge} = -21.1$, $t(7969) = -2.8$, $p < .05$, such that there was a smaller effect of word frequency at the 0 SOA than at the 750 SOA. There was also a significant interaction between SOA and task switch/repetition, $\beta^{\wedge} = 84.3$, $t(7969) = 10.9$, $p < .01$, in which the effect of SOA was smaller when the previous task was the same as the current task, relative to the size of the SOA effect when the previous task was different. There was no significant interaction between task switch/repetition and word frequency. The standard deviation of the random effect of item was estimated at 39.3. The standard deviation of the by-subject adjustments was estimated at 96.2. The residual standard deviation was 170.3.

Reading Aloud Errors. There was a main effect of SOA, $\beta^{\wedge} = -1.1$, $z = -2.0$, $p < .05$ such that more errors were made at the 0 SOA. A significant main effect of previous RT, $\beta^{\wedge} = -.0003$, $z = -2.4$, $p < .05$ suggests that how quickly subjects responded on the previous trial predicts whether they made an error on the current trial; and a main effect of task switch, $\beta^{\wedge} = -.9$, $z =$ -

3.4, p < .01, such that more errors were made when switching between tasks rather than when repeating tasks. No other main effects (word frequency) or interactions approached significance. The standard deviation of the random effect of item was estimated at 1.1. The standard deviation of the by-subject adjustments was estimated at 1.0.

Case Decision RTs. There was no main effect of word frequency, $\beta^{\wedge}$ = 1.0, $t(8090)$= 0.2, p >.05. There was a main effect of SOA, $\beta^{\wedge}$ = 277.1, $t(8090)$= 42.5, p < .01. There was also a main effect of previous RT, $\beta^{\wedge}$ = *0.09*, $t(8090)$ = 14.7, p < .01, and of task switch/repetition, $\beta^{\wedge}$ = *56.4*, $t(8090)$ = 8.7, p < .01. The interaction between SOA and task switch/repetition was significant, $\beta^{\wedge}$ = *76.8*, $t(8090)$ = 8.3, p < .05. There were no other significant interactions. The standard deviation of the random effect of item was estimated at .005. The standard deviation of the by-subject adjustments was estimated at 96.9. The residual standard deviation was 206.6.

Case Decision Errors. There was a main effect of SOA, $\beta^{\wedge}$ = -7.1, z = -4.1, p < .01 and a main effect of trial, $\beta^{\wedge}$ = *3.4*, z = 5.2, p < .01. No other main effects or interactions approached significance. The standard deviation of the random effect of item was estimated at .35. The standard deviation of the by-subject adjustments was estimated at .56.

Vincentiles. The same vincentizing procedure as described previously was used here. As in Experiment 2 it is expected that if SOA and word frequency are generally under-additive (i.e., not just in the means) then the effect of word frequency should be smaller in the 0 SOA condition relative to the 750 SOA condition throughout the distribution. As can be seen in the difference scores (low frequency – high frequency; see bottom panel of Figure 6) this is the case for the reading aloud condition; the effect of word frequency in larger in the 750 SOA condition than in the 0 SOA condition. Again, the effect of word frequency does not center on zero in the 0

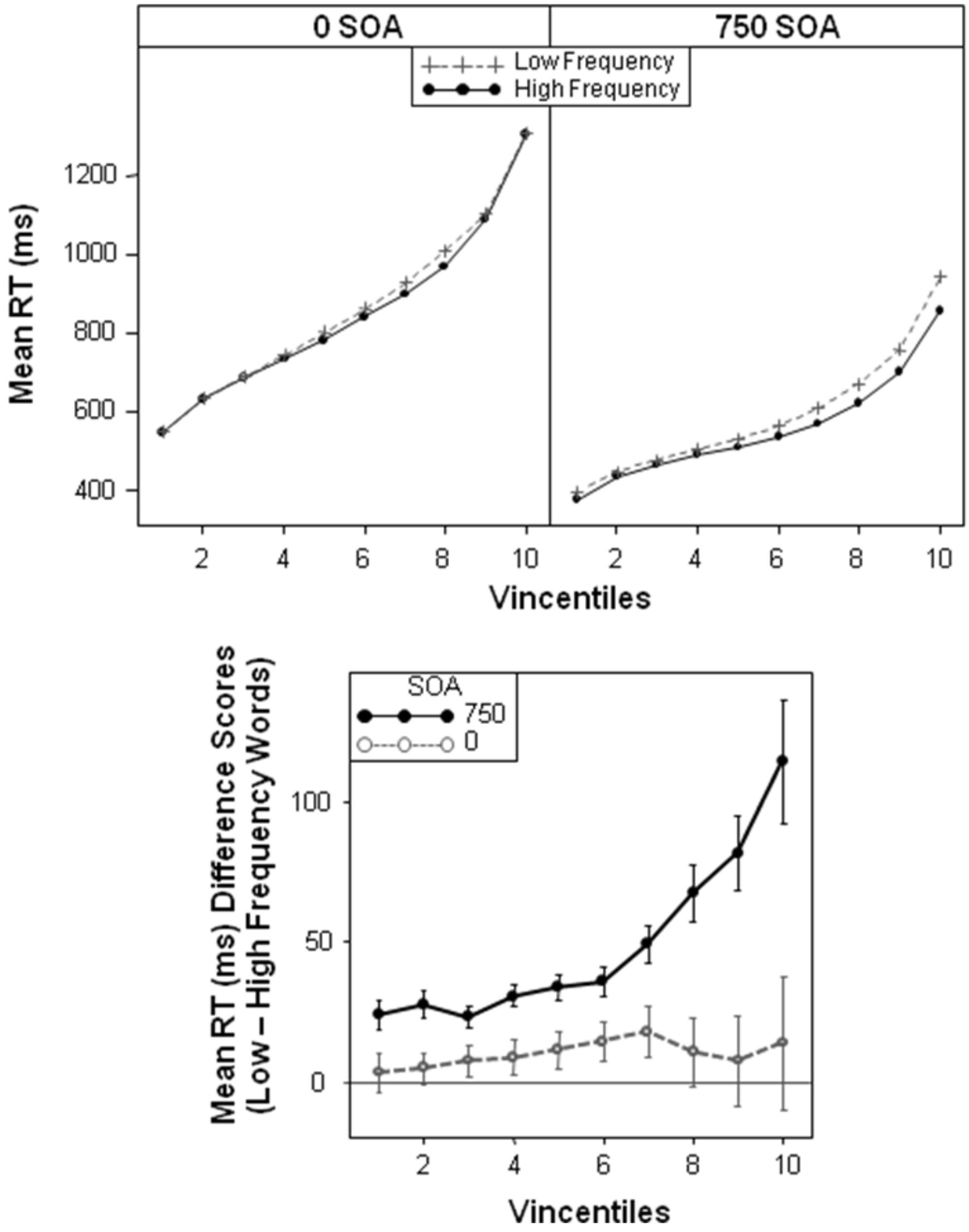


Figure 6. Top panel: Vincentile means for participant’s reading aloud RTs in Experiment 3 as a function of SOA and word frequency. Bottom panel: The difference in vincentile means for low versus high frequency items for participant’s reading aloud RTs. Vertical bars represent the standard error of the mean.

SOA condition as would be expected if the effect of word frequency was completely eliminated. This suggests there is some residual effect of word frequency.

Discussion

Both Experiments 2 and 3 yielded a significant under-additive interaction between SOA and word frequency when reading aloud such that the effect of word frequency is smaller at the 0 SOA than at the 750 SOA. This stands in direct contrast to Experiment 2 of Paulitzki et al. (2009) in which under virtually the same conditions they found no significant interaction between word frequency and SOA. Paulitzki et al. took their findings to imply that lexical processing does not unfold in parallel with cue processing. In contrast, these results show that, to a large extent, lexical processing does occur while the cue is being processed. These results also suggest that early processing (feature and letter analysis) can go on in parallel with cue decoding. This provides some evidence that the additivity observed between nonword complexity and SOA in Experiment 1 is not due to feature and letter level analysis being delayed.

One difference between Experiment 2 and Paulitzki et al. (2009) is that the responses to the case decision task were made vocally rather than via a key press. One might therefore suppose that response mode of the two tasks determines whether lexical processing occurs during the time that the task cue is being processed. However, the results of Experiment 3 suggest that response mode is irrelevant given that this experiment also produced a strong under-additive interaction in which the word frequency effect is smaller at the zero SOA than at the long SOA when using the same response modes as in Paulitzki et al.

Why then, did I observe a significant under-additive interaction between SOA and word frequency whereas Paulitzki and colleagues did not? Several possibilities merit consideration.

First, if the present items are easier to read than the items used by Paulitzki et al., this could explain the different findings. However, this is not a viable explanation given that the present items consisted of monosyllabic and polysyllabic words, ranging from 3-9 letters and, most critically, the low frequency items used here were lower frequency overall than in Paulitzki and colleagues experiment. It cannot, therefore, be argued that the items in the present experiments are easier to read.

Another explanation for the different results seen in the Paulitzki et al. experiment and the ones reported here is that different levels of reading skill are in play. Recent research suggests that reader skill is an important factor in the context of the Psychological Refractory Period (PRP) paradigm. In the PRP paradigm subjects are asked to perform two tasks in a specific order, and the SOA between the stimuli for each task is manipulated. As in the Task Set paradigm, cognitive slack logic is used to explain an underadditive interaction or additive effects between SOA and some manipulated factor in Task 2. In particular, Ruthruff, Allen, Lien and Grabbe (2008) reported that reading skill affects whether SOA and word frequency interact in the context of the PRP paradigm. Skilled readers yielded an under-additive interaction between word frequency and decreasing SOA, whereas less skilled readers yielded additive effects of these factors. If reading skill dictates whether an interaction between SOA and word frequency is observed then readers who have shorter RTs in the 750 SOA condition (i.e., faster reading times, which has been shown to correlate with reading skill) should show an underadditive interaction between SOA and word frequency but subjects with longer RTs in the 750 SOA condition (i.e., slower reading times) should show no interaction between these factors. This is not the case however; when I combined the data across Experiments 2 and 3 and did a median split based on the average reading time at the 750 SOA, both groups of subjects showed a significant underadditive interaction between SOA and word Frequency (for fast subjects the interaction was 17 ms, *F(1, 50)* = 4.1, *MSE* = 848.8, *p* = .05; for slow subjects the interaction was actually larger at 42 ms, *F(1, 50)* =16.2, *MSE* = 1383.7, *p* < .01) despite a very large difference in the average RTs at the 750 SOA (516 ms for fast subjects, 692 ms for slow subjects). Interestingly, the slow readers here were slightly slower than the subjects in Paulitzki et al., who had an average RT of 661 ms at the 750 SOA (note that all experiments used the same lab equipment, so the differences are unlikely to be due to technical issues). This analysis makes it unlikely that the difference in results across the present experiments and that of Paulitzki et al. is due to reading skill. In the absence of any other plausible explanation the most likely conclusion is that Paulitzki et al.'s failure to see an underadditive interaction between SOA and word frequency reflects a Type II error.

The Residual Effect of Word Frequency

Across Experiments 2 and 3 there was a residual effect of word frequency at the 0 SOA, *t*(101) = -2.2, p < .05. If the target word was always *completely* processed while the cue was being decoded no effect of word frequency should be observed (i.e., the difference of the means should be zero with an unknown error variance). However, as can clearly be seen in the vincentiles for both experiments, the effect of word frequency does not center on zero. There are 3 explanations that might be proposed for this: (1) the cognitive slack produced by cue processing is sometimes insufficient to fully absorb the extra time required to process the low frequency words, resulting in the small but significant residual word frequency effect. This account also predicts that on task switching trials, when the effect of SOA is larger (thereby increasing the amount of cognitive slack) the residual effect of word frequency should be smaller than on task repetition trials (when the effect of SOA is smaller, resulting in a smaller amount of cognitive slack). This is not the case; in Experiment 2 the residual effect of word frequency (i.e., the difference between low – high frequency words at the 0 SOA) was 12 ms on task repetition trials and 7 ms on task switch trials. In Experiment 3 the residual effect of word frequency was 15 ms on task repetition trials and 16 ms on task switch trials. Thus, the data do not support this account.

(2) A different account is that on all trials, some word processing is completed while the cue is being processed, but not all of it. This would be true if there is some stage that is affected by word frequency that is not resolved by the time the cue has finished being processed. Support for this account comes from Balota and Abrams (1995) who concluded that there is some effect of word frequency that occurs very late in processing (i.e., during speech production).

(3) Yet another possibility is that on some of the trials word processing occurs in parallel with cue processing, but not on all trials. When this is averaged across trials it would result in a small residual word frequency effect as seen here. This could happen if subjects are mind wandering, momentarily forgetting what the task is, or other minor distractions. Although the first account can be countered by the present data, I cannot distinguish between the second two accounts at present (nor need they be mutually exclusive). Further research is needed to better understand this residual effect of word frequency.

Thus far I have demonstrated that sublexical processing does not occur prior to a task set being in place. In contrast, lexical processing *can* begin prior to a task set being in place. The next experiment addresses the issue of whether lexical processing *always* unfolds while the cue is being processed, or whether context plays a role.

Experiment 4: The Importance of Context

The difference in results between Experiment 1 on the one hand, and Experiments 2 and 3 on the other, suggests a major distinction between sublexical processing, which does not begin until task cue processing has finished, and lexical processing, which can go on in parallel with task cue processing. One account of this difference is that sublexical processing is *not* automatic, whereas lexical processing *is* automatic in the specific sense that sublexical processing is interfered with by cue decoding, whereas lexical processing is not, and therefore does not appear to require intention (in the form of a task set) or "attention" (note that I am not considering spatial attention here).

Experiment 4 addresses a novel question in the context of the Task Set paradigm. Namely, what effect does mixing words and nonwords have on performance? The standard view that lexical processing is automatic makes the straightforward prediction that mixing words and nonwords together in an experiment will yield the same pattern as when words and nonwords are blocked. In particular, the effect of word frequency will still be underadditive with SOA. However, it is also known that subtle changes in the context can have profound effects on performance. In particular, O'Malley and Besner (2008) and Besner, O'Malley and Robidoux (2010) reported a qualitative change in how words are read aloud when they are mixed with nonwords as compared to when only words appear in the experiment. A very different hypothesis, then, is that subjects will adopt a global task set in response to the intermixing of words and nonwords in which they refrain from initiating any target processing (including lexical processing) during cue decoding. To put this another way, given that subjects must wait for cue processing to finish on half of the trials before processing the target (i.e., when the target is a nonword), they may (unconsciously) adopt an experiment wide set in which target processing is delayed until cue decoding is finished on *all* trials. If so, then additive effects of word frequency and SOA are expected, along with additive effects of nonword complexity and SOA. To anticipate the results, this is exactly what is seen.

Method

Subjects. Forty undergraduate students were recruited from the Psychology undergraduate student subject pool at the University of Waterloo. Each subject was awarded credit towards one of their courses for their participation in a single session lasting 35-40 minutes. All subjects reported English as their first language and had normal or corrected-to-normal vision.

Stimuli. The nonword stimuli were the same as those used in Experiment 1; the word stimuli were the same as those used in Experiment 2. As in the previous Experiments the full stimulus set was divided into four: half the items were assigned to the reading aloud task and the other half to the case decision task. Half of the items within each task were assigned to the 0 SOA condition and the other half where assigned to the 750 SOA condition. This assignment of items to tasks and SOA was counterbalanced; subjects were assigned to a stimulus list counterbalance based on order of arrival in the lab.

Design. The design consisted of a 2 x 2 x 4 factorial in which the factors were Task (Reading Aloud vs. Case Decision), SOA (0 vs. 750 ms) and stimulus type (High Frequency vs. Low Frequency, short/basic nonwords vs. long/complex nonwords). All conditions were randomly intermixed within a single block of trials. Each subject received a different random sequence.

Procedure. The procedure, apparatus and task cues were the same as in Experiment 1. Subjects were told they would either see a word or a string of letters. In the reading aloud task they were to pronounce the target (word or nonword).

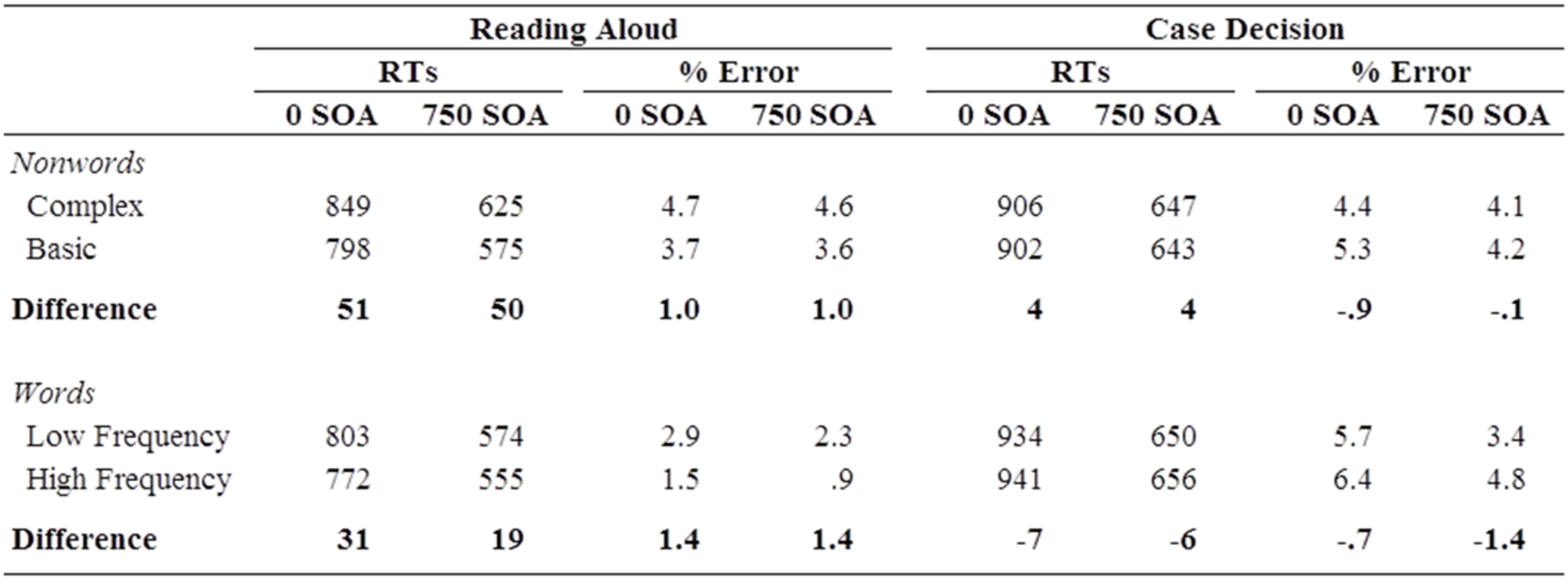
Results

Mean RTs and percentage errors for each condition of Experiment 4 can be seen in Table 4. RTs and errors for the words and for the nonwords were all fitted to different linear mixed-effects models with subject and items as crossed random effects. Responses to each task were fitted to different models. The following factors were included in all initial models: SOA, word frequency, Previous RT (RT on trial N-1), Task switch/repetition (whether the task to be performed on the current trial was the same as the previous trial or different), counterbalance and trial number. The effect of counterbalance was never significant and so was dropped from the final models.

Incorrect responses (3.9%) and spoiled trials (4.4%) were discarded prior to the RT analysis. Outliers were removed using the same trimming procedure as in Experiment 1 resulting in an additional 1.9% of the correct RTs to words and 2.2% of the correct RTs to nonwords being removed in the reading aloud task. 1.8% of the correct RTs to words and 1.8% of the correct RTs to nonwords were removed using the outlier procedure in the Case Decision task.

Reading Aloud Nonwords

RTs. There was a main effect of nonword complexity, $\beta^\wedge$ = 53.1, $t(3836)$ = 4.9, p < .01, and a main effect of SOA, $\beta^\wedge$ = 203.7, $t(3836)$ = 21.1, p < .01. There was also a main effect of previous RT, $\beta^\wedge = 0.09$, $t(3836)$ = 13.0, p < .01, and of trial, $\beta^\wedge$ = -.11, $t(3836)$ = -5.6, p < .01. The effect of task switch/repetition was not significant, $\beta^\wedge = 10.4$, $t(3836)$ = 1.3, p > .05, although it did interact with SOA, such that there was a greater effect of SOA on task switch



	Reading Aloud				Case Decision			
	RTs		% Error		RTs		% Error	
	0 SOA	750 SOA	0 SOA	750 SOA	0 SOA	750 SOA	0 SOA	750 SOA
Nonwords								
Complex	849	625	4.7	4.6	906	647	4.4	4.1
Basic	798	575	3.7	3.6	902	643	5.3	4.2
Difference	**51**	**50**	**1.0**	**1.0**	**4**	**4**	**-.9**	**-.1**
Words								
Low Frequency	803	574	2.9	2.3	934	650	5.7	3.4
High Frequency	772	555	1.5	.9	941	656	6.4	4.8
Difference	**31**	**19**	**1.4**	**1.4**	**-7**	**-6**	**-.7**	**-1.4**

Table 4. Mean Reaction Times (ms) and Percentage Errors (% Errors) in Experiment 4 as a function of Task, SOA, Nonword Complexity and Word Frequency.

trials than on task repetition trials, $\beta^{\wedge} = 42.4$, $t(3836) = 3.7$, $p < .01$. Critically, the interaction between nonword complexity and SOA was not significant $\beta^{\wedge} = 2.5$, $t(3836) = .2$, $p > .05$. The standard deviation of the random effect of item was estimated at 52.9. The standard deviation of the by-subject adjustments was estimated at 158.4. The residual standard deviation was 198.1.

Errors. There was no main effect of SOA, $\beta^{\wedge} = .02$, $z = .09$, $p > .05$, or of nonword complexity, $\beta^{\wedge} = 0.3$, $z = 1.2$, $p > .05$, task switching/repetition, $\beta^{\wedge} = 0.2$, $z = 1.4$, $p > .05$ or trial, $\beta^{\wedge} = 0.0$, $z = .7$, $p > .05$. There was a significant main effect of Previous RT, $\beta^{\wedge} = -0.0$, $z = -2.8$, $p < .01$. None of the interactions approached significance. The standard deviation of the random effect of item was estimated at 1.1. The standard deviation of the by-subject adjustments was estimated at 1.8.

Vincentiles. The same vincentizing procedure described in Experiment 1 was used. If SOA and nonword complexity are generally additive (i.e., not just in the means) then the effect of nonword complexity should be about the same at the 0 SOA condition relative to the 750 SOA condition throughout the distribution. As can be seen in the difference scores (long/complex – short/basic; see bottom panel of Figure 7) this is the case for the reading aloud condition.

Case Decision for Nonwords

RTs. There was no main effect of nonword complexity, $\beta^{\wedge} = 2.6$, $t(5396)= 0.3$, $p > .05$. There was a main effect of SOA, $\beta^{\wedge} = 210.1$, $t(5396) = 24.1$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = .14$, $t(5396) = 19.3$, $p < .01$, trial, $\beta^{\wedge} = -.09$, $t(5396) = -4.1$, $p < .01$ and of task switch/repetition, $\beta^{\wedge} = 88.5$, $t(5396) = 10.1$, $p < .01$. The interaction between SOA and task switch/repetition was significant, $\beta^{\wedge} = 97.3$, $t(5396) = 7.9$, $p < .01$. There were no other significant interactions. The standard deviation of the random effect of item was estimated at

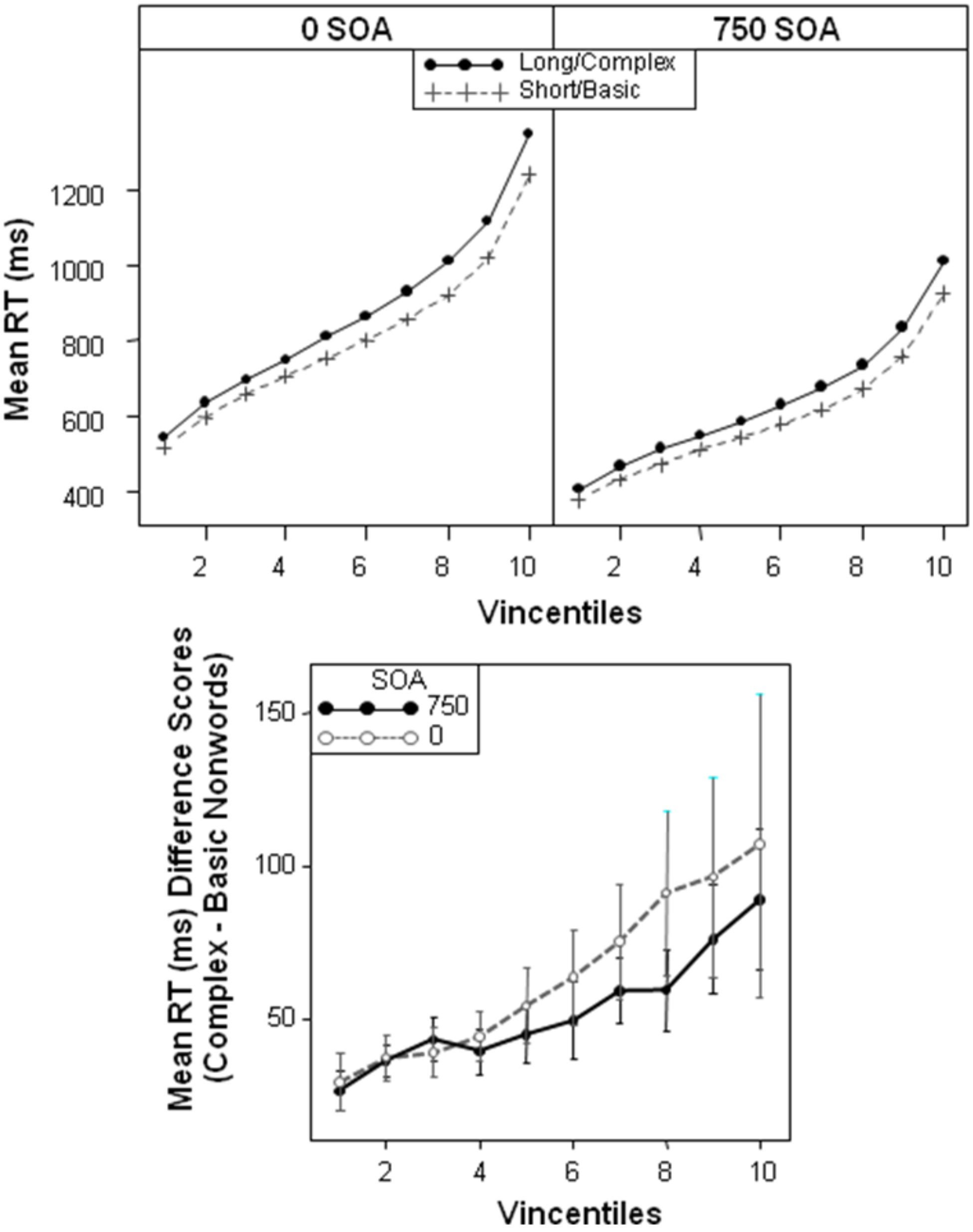


Figure 7. Top panel: Vincentile means for participant's reading aloud RTs in Experiment 4 as a function of SOA and nonword complexity. Bottom panel: The difference in vincentile means for complex versus basic nonwords for participant's reading aloud RTs. Vertical bars represent the standard error of the mean.

24.9. The standard deviation of the by-subject adjustments was estimated at 162.3. The residual standard deviation was 224.8.

Errors. There was no main effect of SOA, $\beta^\wedge = 7.3$, $z = .4$, $p > .05$, nonword complexity, $\beta^\wedge = 2.7$, $z = .14$, $p > .05$, task switch/repetition, $\beta^\wedge = 1.4$, $z = 1.1$, $p > .05$ or trial $\beta^\wedge = 0.0$, $z = .9$, $p > .05$. There was a main effect of previous RT, $\beta^\wedge = -2.9$, $z = -3.3$, $p < .01$. None of the interactions approached significance. The standard deviation of the random effect of item was estimated at .27. The standard deviation of the by-subject adjustments was estimated at .86.

Reading Aloud Words

RTs. There was a main effect of word frequency, $\beta^\wedge = 23.4$, $t(3836) = 2.3$, $p < .05$, and a main effect of SOA, $\beta^\wedge = 191.9$, $t(3836) = 21.6$, $p < .01$. There was also a main effect of previous RT, $\beta^\wedge = 0.05$, $t(3836) = 8.3$, $p < .01$, and of trial, $\beta^\wedge = -.07$, $t(3836) = -2.5$, $p < .05$. The effect of task switch/repetition was not significant, $\beta^\wedge = 13.5$, $t(3836) = 1.5$, $p > .05$, although it did interact with SOA, such that there was a greater effect of SOA on task switch trials than on task repetition trials, $\beta^\wedge = 56.7$, $t(3836) = 4.4$, $p < .01$. Critically, there was no interaction between word frequency and SOA, $\beta^\wedge = 4.4$, $t(3836) = .3$, $p > .05$. The standard deviation of the random effect of item was estimated at 35.8. The standard deviation of the by-subject adjustments was estimated at 135.4. The residual standard deviation was 197.7.

Errors. There was no main effect of SOA, $\beta^\wedge = .6$, $z = 1.4$, $p > .05$, previous RT, $\beta^\wedge = 0.0$, $z = .4$, $p > .05$ or trial, $\beta^\wedge = 0.0$, $z = .6$, $p > .05$. There was a significant main effect of word frequency, $\beta^\wedge = 1.1$, $z = 2.1$, $p < .05$ and a main effect of task switch/repetition, $\beta^\wedge = 1.1$, $z = 4.1$, $p < .01$, such that more errors were made when switching between tasks rather than when repeating tasks. None of the interactions approached significance. The standard deviation of the random effect of item was estimated at 2.2. The standard deviation of the by-subject adjustments was estimated at 1.9.

Vincentiles. Again, if SOA and word frequency are generally additive then the effect of word frequency should be the same at both levels of throughout the distribution. As can be seen in the difference scores (low frequency – high frequency; see bottom panel of Figure 8) this is the case. Although it appears that the effect of word frequency does not increase across vincentiles, when looking at the standard error bars it is clear that there is significant overlap between this and the vincentiles at the 750 SOA. This suggests that the two distributions overlap, as expected by the additivity seen in the mean RTs.

Case Decision for Words

RTs. There was no main effect of word frequency, $\beta^{\wedge} = 4.4$, $t(3876) = 0.5$, $p > .05$. There was a main effect of SOA, $\beta^{\wedge} = 219.7$, $t(3876) = 20.3$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = 0.12$, $t(3876) = 11.4$, $p < .01$, trial, $\beta^{\wedge} = -.06$, $t(3876) = -2.1$, $p < .05$ and of task switch/repetition, $\beta^{\wedge} = 88.5$, $t(3876) = 8.1$, $p < .01$. SOA and task switch/repetition interacted such that there was a greater effect of task switch at the 0 SOA than at the 750 SOA, $\beta^{\wedge} = 113$, $t(3876) = 7.3$, $p < .01$. There were no other significant interactions. The standard deviation of the random effect of item was estimated at .23.4. The standard deviation of the by-subject adjustments was estimated at 179.7. The residual standard deviation was 237.5.

Errors. There was a main effect of Trial, $\beta^{\wedge} = 0.02$, $z = 2.9$, $p < .01$. No other main effects or interactions approached significance. The standard deviation of the random effect of item was estimated at .43. The standard deviation of the by-subject adjustments was estimated at .67.

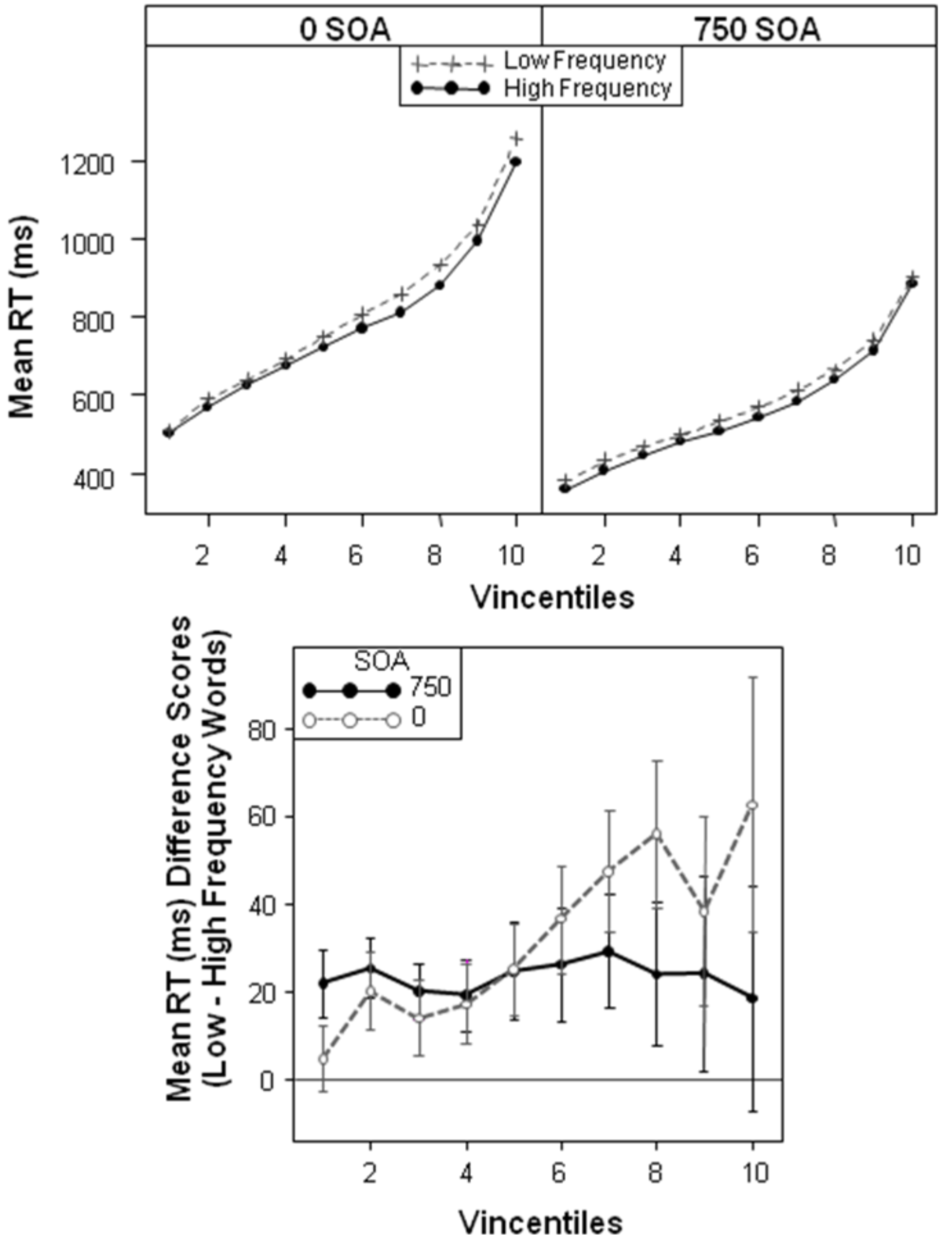


Figure 8. Top panel: Vincentile means for participant's reading aloud RTs in Experiment 4 as a function of SOA and word frequency. Bottom panel: The difference in vincentile means for low versus high frequency items for participant's reading aloud RTs. Vertical bars represent the standard error of the mean.

Discussion

As in Experiment 1, there was no interaction between nonword complexity and SOA; this is consistent with the claim that sublexical processing waits until the cue has been decoded before it can begin. The novel result, however, is that when reading *words* aloud, there was no interaction between word frequency and SOA, not even a trend towards underadditivity. This stands in direct contrast to Experiments 2 and 3 in which an underadditive interaction between these two factors was observed (using the same stimuli, procedure and apparatus). The only difference is that in the present experiment nonwords were randomly intermixed whereas in Experiments 2 and 3, only words were presented. Thus, yet again, the nature of the stimulus set affects how subject's performance unfolds (see O'Malley & Besner, 2008; Besner et al, 2010).

Given that subjects in the present experiment must, on half the trials, wait until cue decoding has finished in order to begin sublexical processing of nonword targets, the simplest account is that an experiment-wide set is adopted in which subjects always wait until the cue has been processed before initiating target processing. The observation that word frequency and SOA have additive effects when words and nonwords are mixed together implies that lexical processing *is* delayed, and consequently processing cannot be construed as automatic because according to that view such processing *cannot* be delayed. Instead, subjects are able to control (presumably unconsciously) whether or not they engage in lexical processing during the time they are decoding the cue.

More generally, the present results support the idea that context is a critical determinant of how processing unfolds over time. To date, the word recognition literature has in large measure failed to embrace the idea that context strongly affects how processing takes place, perhaps because it is overly enamored of the idea that processing is "automatic" in a variety of ways.

Alternative accounts

Sublexical Processing. It could be that the presence of nonwords simply encourages subjects to read sublexically on *all* trials; many of the words used in the stimulus set are regular (that is, they follow the standard grapheme to phoneme conversion rules) and so are able to be pronounced correctly via the sublexical route. This possibility has been the subject of some debate between Zielger, Perry and Zorzi (2009) and Besner and O'Malley (2009). However, there is a significant word frequency effect at both SOA's in both Experiment 2 and 3. This suggests that subjects are not adopting a strategy in which they emphasize sublexical processing since such processing is insensitive to the effect of word frequency.

Serial Processing. Another account of the present results is that subjects process target and cue sequentially, but process the target prior to processing the task cue at the 0 SOA. This eliminates the cognitive slack generated by processing target and cue simultaneously, and, on its own, is consistent with the observation of no interaction between word frequency and SOA, and nonword complexity and SOA. However, further consideration suggests that this account is also wanting. If subjects process the cue after they process the target, then they must generate responses to the reading task and the case decision task on every trial and hold these responses in abeyance until the cue has been decoded.[6] If the two tasks are performed one after the other

---

[6] A variant of this account is that subjects process one task on every trial, before decoding the cue. If cue processing indicates that they have done the right task then they make a response. If not, they back up and do the other task and then emit a response. Assuming they do the wrong task approximately 50% of the time, an effect of word frequency should be seen on approximately half the word trials for the case decision task at the zero SOA. This would result in a significant word frequency effect on the case decision trials, which is not observed here. The same logic applies to the effect of nonword complexity; on this account there should be an effect

(regardless of the order) then the effect of word frequency or nonword complexity should be seen in the case decision RTs as well in the reading aloud RTs. Given that there is no evidence for either of these effects in the case decision task, this account is not viable. Note that this argument is also applicable to the effect of nonword complexity in Experiment 1.

Parallel Processing. Rather than process the two tasks serially (and then decode the tone) another scenario is that subjects process the target in regards to both tasks in parallel first, and then decode the cue. In this scenario the total time taken to generate both responses (i.e., the total time spent on target processing) is determined by the slower of the two tasks. As can be seen in Table 4, at the 750 SOA, RTs to the case decision task are slower than RTs to the reading aloud task (on average 66 ms slower). At the 0 SOA, if both tasks are being processed in parallel (prior to cue processing) and the case decision task takes longer to perform, then the effect of the manipulated factor should be absorbed into the time taken to process the case decision task. Put another way, for this account to be true, there should be no effect of word frequency or nonword complexity on 0 SOA trials even when reading aloud. Obviously, this is not the case. On the other hand, it might be argued that the case decision task is slower than the reading aloud task because of a response execution component and that the reading aloud task actually takes more time than the case decision task prior to an overt response. In this case both tasks should show a word frequency effect and a nonword complexity effect at the zero SOA. Clearly, the case decision task does not show either of these effects.

Therefore, the best account of the present data is that subjects are processing the cue prior to target processing. Consequently, the fact that word frequency and nonword complexity are

---

of nonword complexity at the zero SOA in the case decision task when the stimulus consists of a nonword. No such evidence is seen in the data.

additive with SOA implies that subjects do not begin to process the target until they have finished processing the cue telling them which task they are to do.

When reading nonwords aloud, factors that directly affect sublexical processing (here, letter length and complexity), are additive with SOA, suggesting that sublexical processing does not begin until the cue has been processed. Further, word frequency, a factor that indexes lexical processing, is also additive with SOA provided that nonwords and words are intermixed in the experiment. This suggests that, in the present context, lexical processing does not begin until cue processing has been completed. In short, the process of reading aloud is affected by intention, that is, the need first to decode the task cue and then implement the correct task set. This counters the claim that reading aloud is automatic in the specific sense noted above, namely, that it always occurs without intention. Taken together with the fact that lexical processing can be carried out in parallel with cue decoding, *provided that only words appear in the experiment*, the clear inference is that reading aloud is strongly context dependent in a way not considered to date by any current theory of reading aloud or by any theory of automaticity.

Thus far I have reported that sublexical processing, as indexed by the effect of nonword length and complexity does not go on in parallel with cue processing, and lexical processing, as indexed by the effect of word frequency, can go on in parallel with cue decoding, depending on the context. Experiment 5 considers whether it is possible to carry out a form of semantic processing during cue processing.

Experiment 5: Semantic Processing

Unlike in the previous experiments, selecting a factor to manipulate that indexes semantic processing poses a bit of a challenge. One possibility is to manipulate imageability. Subjects are faster to read aloud high imageable words (e.g., cat) than low imageable ones (e.g., vice), and this effect is widely thought to reflect semantic processing (Woolams, 2005; Evans, Lambon & Woolams, 2011; Strain, Patterson & Seidenberg, 1995). Problematically, the effect is quite small (often only 10 ms) likely making it difficult to detect an interaction with SOA in the context of the Task Set paradigm. A more robust factor is the number of features of concrete items (i.e., semantic richness); this has also been shown to affect RT. However, there is some debate in the literature as to whether these effects arise during semantic processing or during response selection (Grondin, Lupker & McRae, 2009). If the effect of semantic richness arises during response selection than it cannot be used to index the role of intention on semantic processing, because semantic processing would occur *prior* to the stage at which this factor has its effect.

Beyond selecting a factor to manipulate, assessing whether semantic processing can go on in parallel with cue decoding presents an issue not seen in the previous experiments. That is, activation of semantics only occurs *after* some activation arises in the orthographic input lexicon. If orthographic lexical processing is delayed, then semantic processing will also be delayed, simply because it occurs *after* orthographic lexical processing. So additivity of some semantic factor and SOA would say nothing about whether semantic processing can occur prior to a task set being in place. In order to make the claim that semantic processing does not begin until a task set is in place, it is necessary to show an underadditive interaction between some lexical factor and SOA. Then I can assess whether semantic processing is delayed or not. Thus, I need to index both orthographic and semantic processing in the same experiment. To index orthographic/

phonological lexical processing, I manipulated word frequency in the reading aloud task. To index semantic processing, I used antonym generation as the second task (subjects were asked to give the antonym of the word they were presented with), which can only be done by recourse to semantics.

Within antonym generation I also need to manipulate some factor that affects semantic processing (so that I can measure whether this factor interacts with SOA or not). In Experiments 2 and 3, I manipulated word frequency because it is generally accepted that this factor affects lexical processing. In the context of the DRC model, it is considered to arise within the OIL and the POL. According to PDP models, word frequency arises from the strength of connections between levels (orthography to phonology *and* between orthography to semantics, and semantics to phonology). As well, Besner and colleagues argue, within a localist framework, that frequency affects the strength of connections between the OIL and the POL (lexical processing) *and* connections between OIL and semantics, and semantics and the POL (McCann & Besner, 1987; Besner & Smith 1992; Borowksy & Besner 1993; Blais, O'Malley & Besner, 2011; Besner, Moroz & O'Malley, 2011). As can be seen in Figure 9, this account argues that word frequency affects routes A, B and C. Given this, in Experiment 5, I manipulated word frequency in *both tasks* which allows me to determine whether lexical processing goes on in parallel with cue processing (via the reading aloud task), and whether semantic processing goes on in parallel with cue processing (via the antonym generation task).

Several outcomes merit consideration. When reading aloud, the effect of word frequency may go under-additive with SOA as seen in Experiments 2 and 3; suggesting that reading aloud occurs in parallel with cue processing. Alternatively, the experimental context might affect the outcome as seen in Experiment 4; lexical processing might also be delayed (word frequency

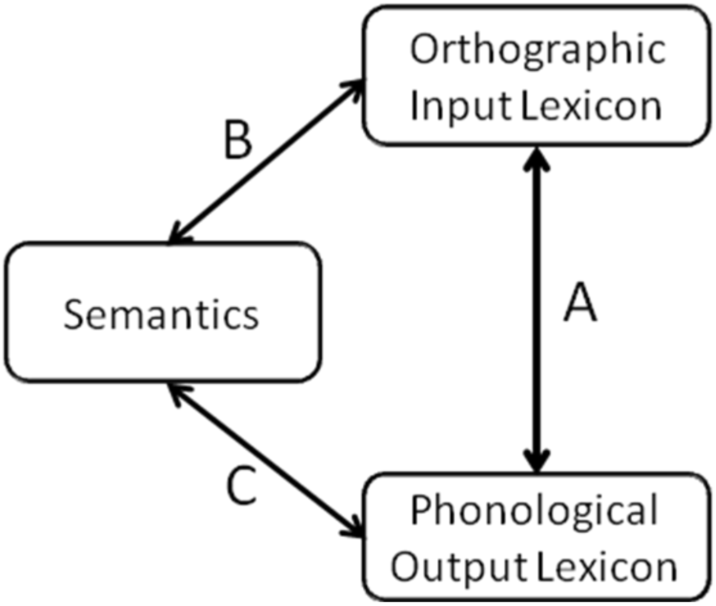


Figure 9. A schematic of lexical and semantic processing in the DRC model. The strength of connections between each module (marked by A, B and C) may be affected by word frequency.

is additive with SOA when reading aloud). If SOA and word frequency are additive in reading aloud task, then no conclusions can be drawn about semantic processing, regardless of the outcome in the antonym generation task.

In contrast, if lexical processing does go on in parallel with cue processing (i.e., there is an interaction between SOA and word frequency in the reading aloud task), then I can use the results from the antonym generation task to determine whether semantic processing can occur prior to a task set being in place or not. In the antonym generation task, if there is an underadditive interaction between word frequency and SOA this suggests that semantic processing can occur prior to a task set being in place; if word frequency and SOA are additive in the antonym task this suggests that some aspect of semantic processing does not begin until a task set is in place.

Method

Subjects. Forty-eight undergraduate students were recruited from the Psychology undergraduate student subject pool at the University of Waterloo. Each subject was awarded credit towards one of their courses for their participation in a single session lasting 25-30 minutes. All subjects reported English as their first language and had normal or corrected-to-normal vision.

Design. The design was a 2 x 2 x 2 factorial in which the factors were Task (Reading Aloud vs. Antonym Generation), SOA (0 vs. 750 ms) and word frequency (high vs. low). All conditions were randomly intermixed within a single block of trials. Each subject received a different random sequence.

Procedure. Procedure, apparatus and stimuli were the same as Experiment 2; the only difference was in the instructions for the second task. Subjects were told to say the first antonym they could think of in response to the target. An error was recorded if the response was not an antonym of the target word. The most common errors were synonyms of the target word or a response of “not-‘target word’” (e.g., in response to “always” people might say “forever” or “not-always”; both were recorded as errors). An error was also recorded if the subject couldn’t generate a response in a reasonable amount of time (over approximately 10 seconds). In Appendices B and C the antonyms used to calculate the strength of association between that word and its antonym appears in brackets. Note if subjects generated a different antonym to the one listed there it was marked as correct (e.g., for word “end”, the antonym listed is “begin”, however “start” would also be accepted as correct).

Results

Mean RTs and percentage errors for each condition of Experiment 5 can be seen in Table 5. RTs and errors were fitted to linear mixed-effects models, with subject and items as crossed random effects. Responses to each task were fitted to different models. The following factors were included in all initial models: SOA, word frequency, Previous RT (RT on trial N-1), Task switch/repetition, counterbalance and trial number.

Incorrect responses (8.8%) and spoiled trials (22.0%) were discarded prior to the RT analysis. The number of spoiled trials is larger here than in the previous experiments simply because the longer subjects take to utter a response the more likely something else (e.g., breathing, movement, coughing, stuttering etc.) is to trigger the microphone, resulting in a spoiled trial. The RTs were first fitted to a linear mixed effect model that contained only the main effects of each factor. RTs that were greater than 2.5 standard deviations from the predicted RTs of the model were removed prior to further analysis resulting in an additional 2.6% of the

	Reading Aloud				Antonym Generation			
	RTs		% Error		RTs		% Error	
	0 SOA	**750 SOA**	**0 SOA**	**750 SOA**	**0 SOA**	**750 SOA**	**0 SOA**	**750 SOA**
Low frequency	878	661	2.1	1.2	1990	1700	25.3	25.1
High Frequency	853	612	1.4	.9	1581	1337	12.9	11.2
Difference	**25**	**49**	**.7**	**.3**	**409**	**363**	**12.4**	**13.9**

Table 5. Mean Reaction Times (ms) and Percentage Errors (% Errors) in Experiment 5 as a function of Task, SOA, and Word Frequency.

correct RTs being removed from the Reading Aloud task and 2.5% of the correct RTs being removed from the Antonym Generation task.

Reading Aloud RTs. There was a main effect of word frequency, $\beta^{\wedge} = 52.5$, $t(6399) = 6.2$, $p < .01$, and a main effect of SOA, $\beta^{\wedge} = 192.5$, $t(6399) = 23.2$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = 0.02$, $t(6399) = 8.9$, $p < .01$, and of task switch/repetition, $\beta^{\wedge} = 58.3$, $t(6399) = 8.1$, $p < .01$. Task switch/repetition interacted with SOA such that the effect of SOA was greater on task switch trials that on task repetition trials $\beta^{\wedge} = 88.3$, $t(6399) = 9.1$, $p < .01$. Critically, the interaction between word frequency and SOA was significant, $\beta^{\wedge} = 21.6$, $t(6399) = 2.2$, $p < .05$; the effect of word frequency was the smaller at the 0 SOA than at the 750 SOA. The standard deviation of the random effect of item was estimated at 44.6. The standard deviation of the by-subject adjustments was estimated at 269.5. The residual standard deviation was 107.3.

Reading Aloud Errors. There was no main effect of word frequency, $\beta^{\wedge} = .3$, $z = .4$, $p > .05$, or SOA, $\beta^{\wedge} = .5$, $z = 1.5$, $p > .05$. There were no main effects of counterbalance, previous RT or trial. There was a significant main effect of task switch/repetition, $\beta^{\wedge} = .9$, $z = -3.6$, $p < .01$. There was no significant interaction between SOA and word frequency, $\beta^{\wedge} = .003$, $z = .007$, $p > .05$. The standard deviation of the random effect of item was estimated at 3.7. The standard deviation of the by-subject adjustments was estimated at 1.8.

Vincentiles. If SOA and word frequency are generally under-additive (i.e., not just in the means) then the effect of word frequency should be smaller in the 0 SOA condition relative to the 750 SOA condition throughout the distribution. The difference scores (low frequency – high frequency; see bottom panel of Figure 10) for the reading aloud condition show that this is the case, and that the effect of word frequency increases with increasing vincentiles in the 750 SOA condition, but not in the 0 SOA condition. As in Experiments 2 and 3, the effect of word

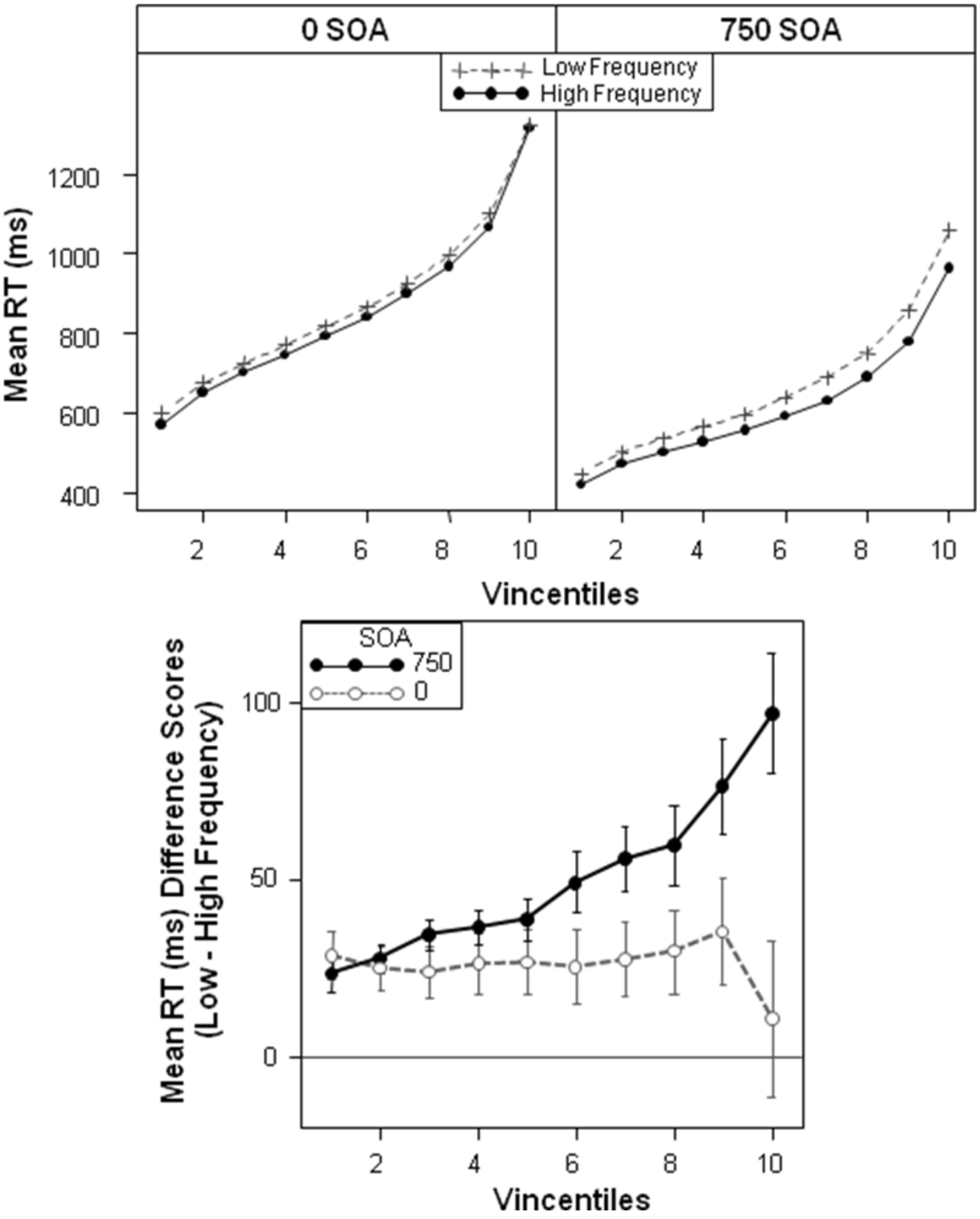


Figure 10. Top panel: Vincentile means for participant's reading aloud RTs in Experiment 5 as a function of SOA and word frequency. Bottom panel: The difference in vincentile means for low versus high frequency items for participant's reading aloud RTs. Vertical bars represent the standard error of the mean.

frequency does not center on zero in the 0 SOA condition, suggesting that there is some residual effect of word frequency (see discussion section relevant to Experiments 2 & 3).

Antonym Generation RTs. There was a main effect of word frequency, $\beta^{\wedge} = 397.1$, $t(4978) = 6.0$, $p < .01$, and a main effect of SOA, $\beta^{\wedge} = 232.3$, $t(4978) = 7.6$, $p < .01$. There was also a main effect of previous RT, $\beta^{\wedge} = 0.05$, $t(4978) = 6.6$, $p < .01$, and of task switch/repetition, $\beta^{\wedge} = 72.9$, $t(4978) = 2.6$, $p < .05$, though it did not interact with SOA, $\beta^{\wedge} = 18.9$, $t(4978) = .5$, $p > .05$. In contrast to the reading aloud task, the interaction between word frequency and SOA was not significant, $\beta^{\wedge} = 36.6$, $t(4978) = 1.0$, $p > .05$. Indeed, there was an over-additive trend of 46 ms, such that the effect of word frequency was (non-significantly) larger at the 0 SOA than at the 750 SOA. The standard deviation of the random effect of item was estimated at 511.5. The standard deviation of the by-subject adjustments was estimated at 241.4. The residual standard deviation was 634.9.

Antonym Generation Errors. There a main effect of word frequency, $\beta^{\wedge} = 1.3$, $z = 5.4$, $p < .01$, task switch/repetition, $\beta^{\wedge} = .2$, $z = 2.6$, $p < .05$, and of trial, $\beta^{\wedge} = .002$, $z = 3.3$, $p < .01$. There were no significant effects of SOA, counterbalance or previous trial. There was no significant interaction between SOA and word frequency, $\beta^{\wedge} = .15$, $z = .9$, $p > .05$. The standard deviation of the random effect of item was estimated at 1.7. The standard deviation of the by-subject adjustments was estimated at .7.

Vincentiles. The vincentile distribution of the difference scores (low frequency – high frequency; see bottom of Figure 11) for the Antonym Generation task show a very different pattern than those for the reading aloud task. Here it’s clear that here the effect of word frequency is the same at both SOA’s, throughout the distribution (i.e., the distributions overlap).

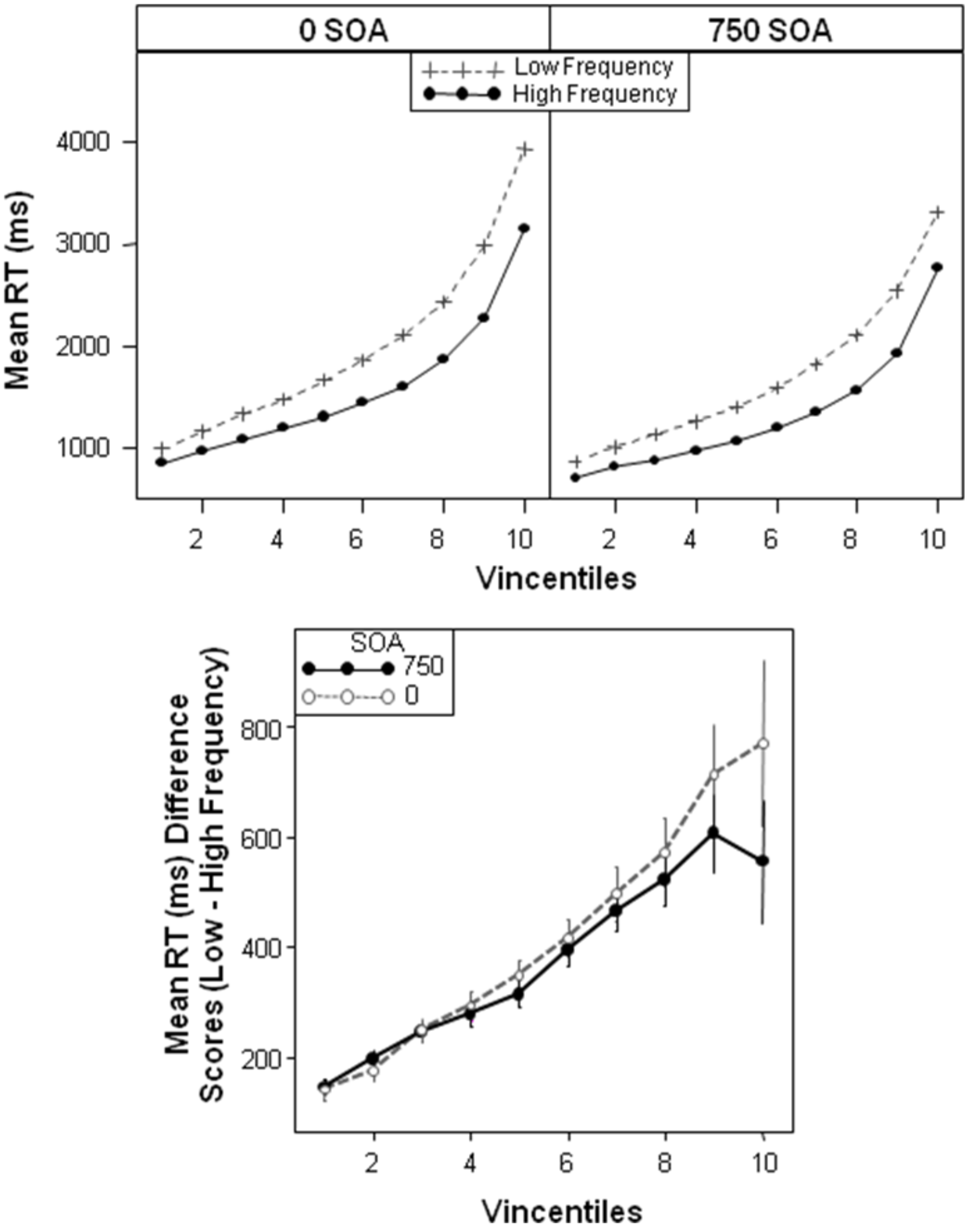


Figure 11. Top panel: Vincentile means for participant's antonym generation RTs in Experiment 5 as a function of SOA and word frequency. Bottom panel: The difference in vincentile means for low versus high frequency items for participant's antonym generation RTs. Vertical bars represent the standard error of the mean.

Discussion

Consistent with the results of Experiments 2 and 3, the effect of word frequency when reading aloud was underadditive with SOA, suggesting that orthographic and phonological lexical processing can go on in parallel with cue processing. The novel result is the additivity between these same two factors in the antonym generation task. This suggests that some aspect of semantic processing does not begin until after cue processing has finished.

Why, under the present conditions, is semantic processing delayed but lexical processing is not? In both tasks used here (reading aloud and antonym generation), the word must be read to produce a correct response. Thus subjects may adopt a default set in which the word is lexically processed, in parallel with cue decoding, on all trials, but the processes involved in antonym generation are only initiated after the cue has been decoded. Put another way, all words are processed via the lexical route (i.e., orthographically and phonologically) but not via the semantic route, prior to a task set being in place. When the task is to read aloud, the effects of word frequency and SOA are underadditive, because orthographic and phonological processing occurs while the cue is being processed. The effects of word frequency and SOA are additive in the antonym generation because semantic processing is initiated after the cue has been processed, and only if the cue indicates the antonym generation task.

Although Besner and colleagues argue that word frequency affects the strength of connections between all modules (paths A, B and C in Figure 9), in DRC word frequency *only* affects processing in the OIL and the POL (path A). However, the results of Experiment 5 present a challenge for this view. If orthographic processing can go on in parallel with cue decoding before a task set is in place and this is the only locus of word frequency effects, then an underadditive interaction between word frequency and SOA should be observed in *both* tasks. In the antonym generation task there should be at least partial underadditivity between SOA and word frequency, but there is no evidence for this (in fact the trend is in the over-additive direction). The simplest explanation is that word frequency affects the strength of connections along the semantic route (paths B and C) as well those in the lexical route (path A). Subjects begin *lexically processing* the word prior to a task set being in place on all trials (or at least on most trials) resulting in the underadditive interaction between word frequency and SOA when reading aloud. Activation is not passed on to semantics (i.e., path B is delayed) until the cue has been processed; thus any effect of word frequency that affects the processes involved specifically with antonym generation will produce additivity between word frequency and SOA.

Consistent with this hypothesis, the effect of word frequency is much larger in the antonym generation task than in the reading aloud task (at the 750 SOA the effect of word frequency is 314 ms larger in the antonym generation task than in the reading aloud task). A component of this difference may arise from scaling differences; the antonym generation task is considerably slower, and one might suppose that much slower tasks yield much larger effects. However, at the long SOA, when the RTs were transformed to the same scale (the z-distribution) the effect of word frequency was still larger in the antonym generation task than in the reading aloud task, $F = 30.6$, $MSE = .022$, $p < .001$. This suggests that scale does not account for *all* of the difference between tasks (though it may account for some of the difference).

The nature of the antonym generation task is quite different from the reading aloud task and can also explain at least some of the difference in the size of the word frequency effect; it requires the subject to retrieve the meaning of one word (e.g., “hate”) and from that to generate a word that means the opposite (e.g., “love”). This requires far more steps than simply reading aloud, many of which are likely to be affected by word frequency. An additional note is that often the antonym of a high frequency word is another high frequency word, and the antonym of a low frequency word is another low frequency word (e.g., “big-small” vs. “victory-defeat”). The average log HAL frequency is 8.9 for the antonyms of the low frequency words, and is 10.5 for antonyms of the high frequency words. This compounds the frequency effect. Overall then, it is clear that word frequency affects more than just lexical processing, and that at least some process associated with antonym generation is delayed until cue processing has finished, in contrast to reading aloud.

General Discussion

The results of the experiments reported here demonstrate that whether sublexical/lexical/semantic processing occurs in parallel with processing a cue that indicates what task to perform depends on the nature of the processing that is being carried out. Sublexical processing does not begin until the cue has been processed (Experiment 1), as does some aspect of semantic processing (Experiment 5). In contrast, lexical processing can begin prior to the task set being in place (Experiments 2 & 3) provided that only words appear in the experiment, but if nonwords are intermixed then even lexical processing waits for cue decoding before it begins (Experiment 4). These results provide strong evidence against the wide spread view that reading aloud is automatic in the sense that domain-specific processing cannot be interfered with and is triggered by the onset of the stimulus.

Alternate Models of Visual Word Recognition

Thus far I have used the dual-route architecture (e.g., DRC) as a framework to discuss my predictions and results. However, parallel distributed processing (PDP) models are also commonly used to explain how people read aloud (Plaut et al., 1996; Seidenberg, 2005). Unlike localist models, PDP models use distributed representations of orthography, phonology and semantics. Hidden layers connect these domain specific layers to each other (See Figure 12). When a word is presented to the model activation of all the units in the model that correspond to that input occurs. That activation then cascades forward to the output units that the model associates with that input. Activation flows both forward and backward, in an interactive fashion throughout the layers of the network (at least in most PDP models, see Plaut et al. 1996 for an example of PDP models that only have feed-forward processing). The connections within and between layers are weighted in response to learning. Reading aloud can be accomplished via

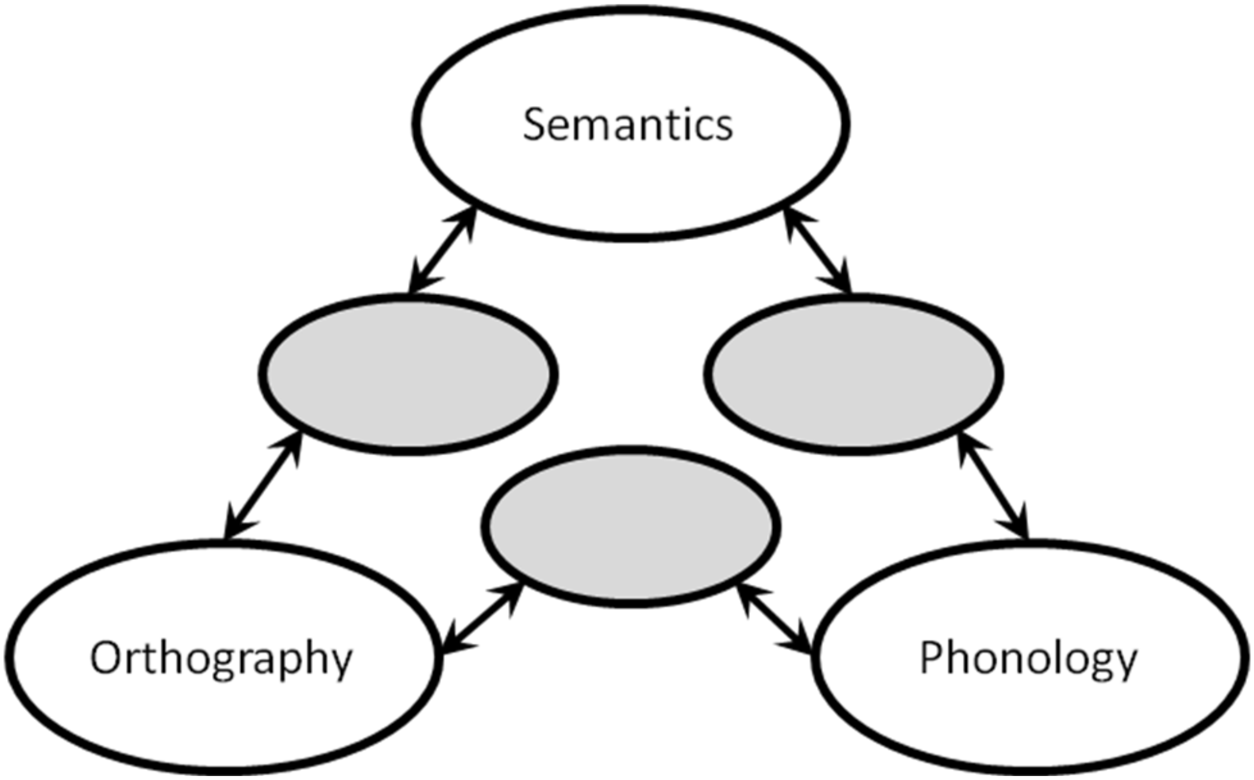


Figure 12. A schematic depiction of a typical PDP model of reading aloud. Large ovals represent groups (or layers) of units that process specific types of information (orthography, phonology and semantics). Grey ovals represent hidden units.

orthography to semantics to phonology; or by orthography to phonology. A known word can be read via either route. An unknown word (i.e., a nonword) can only be read via orthography to phonology as it does not have a semantic association. These models are able to correctly pronounce thousands of words, and many nonwords (i.e., words the model has never read before; see Seidenberg, 2005).

How might one interpret the present results using the PDP framework? Given that nonword processing does not begin until cue processing has finished (Experiment 1), I assume that activation from orthography to phonology is blocked (as this is the only route by which nonwords can be read). Activation from orthography to semantics to phonology must be able to unfold while cue processing is occurring because words can be processed while the cue is being processed (Experiments 2 & 3). Thus, the "orthography to semantics to phonology" route is able to unfold in parallel with cue processing, but not the "orthography to phonology" route. On the surface the results of Experiment 5 (antonym generation) appear to be problematic for this account. The fact that word frequency is additive with SOA when generating an antonym suggests that at some stage semantic processing waits until a task set is in place. However this delay may be at a stage later than semantic activation per se (i.e., while selecting the appropriate antonym). As such, PDP models can be used as a framework to explain much of the present results.

Like the DRC model, PDP models fail in explaining why, when words and nonwords are intermixed, there is no underadditive interaction between SOA and word frequency (Experiment 4). If orthography to phonology is blocked by task set or capacity limitations, words should still be able to be read by orthography to semantics to phonology (and thus should produce an underadditive interaction between word frequency and SOA). One might argue that the connections between layers are differentially affected by task set (i.e., strong connections would not require a task set to be in place but weak connections would). However this is entirely post-hoc. A more reasonable assumption is that in this specific context, both routes are delayed until cue decoding has completed. At present neither PDP models nor the DRC model provide any explanation for how a processing route might be delayed in one context but not in another. Thus, at present, they cannot explain how changing the stimulus set so as to include nonwords (as in Experiment 4) resulted in a quantitatively different pattern of results than when only words were present.

PDP models and localist models differ in how they explain the effect of word frequency. As mentioned above, the connections between distributed nodes (and hidden layers) are weighted; words that occur more often have stronger connections between and within layers, and thus are read faster. The DRC model assumes that the effect of word frequency arises from different resting levels of activation within the localist nodes in orthography and phonology. However, as mentioned previously, Besner and colleagues (McCann & Besner, 1987; Besner & Smith 1992; Borowksy & Besner 1993; Blais et al., 2011; Besner et al., 2011) have proposed that a localist model be retained, but the effect of word frequency arises from differing connection strengths between the OIL and the POL, between the OIL and semantics, and between semantics and the POL (see figure 1).

If I adopt this modified localist model to explain the present results, much of my previous conclusions remain the same, but there are two key differences. First in Experiment 4, where I observed no interaction between word frequency and SOA, processing may have reached the OIL, but the information did not activate the POL until a task set was in place (given that in this account the effect of word frequency resides in the *connections* between OIL and POL).

The other difference relates to Experiment 5 which examined the role of task set on semantic processing. If word frequency affects the connections between the OIL and semantics, then the additivity between word frequency and SOA when generating an antonym suggests activation did not reach semantics until a task set was in place. In the account proposed by Besner and colleagues there are two loci for the effects of frequency within the semantic route (1) the connections between the OIL and semantics (2) the connections between semantics and the POL. This is compared to just one locus within the lexical route: the connections between the OIL and the POL. This may provide some explanation for the larger frequency effects observed in the antonym generation task relative to the reading aloud task; although given the size of the frequency effects, is unlikely to be the whole story. Overall, the modified localist model is comparable to the regular DRC model in terms of its ability to explain the present results.

Intention versus Attention

The Task Set paradigm is closely related to the psychological refractory period (PRP) paradigm, which has been used to assess the role of central attention when reading aloud (e.g., McCann, Remmington & Van Selst, 2000; Reynolds and Besner, 2006; O'Malley, Reynolds, Stolz & Besner, 2008; Besner, O'Malley & Reynolds, 2009). In the PRP paradigm subjects are asked to perform two tasks and make two serially ordered responses. The time between the presentation of the first and second target (SOA) is manipulated such that on some trials there is overlap in the presentation of the stimuli for each task (short SOA) and on some trials the target for Task 1 is presented well before the target for Task 2 (long SOA). As in the task set paradigm the cognitive slack logic outlined in the introduction is typically used to make predictions about performance. The key difference between PRP and the Task Set paradigm is that in the PRP paradigm subjects always know what two tasks they are to perform (and what order to perform them in). Thus, when there is temporal overlap in the presentation of the stimuli, any delay in processing the second target is not due to task set implementation, but instead is due to Task 1 processing, which requires some limited capacity resource, commonly referred to as central attention, which is needed to execute both tasks (Pashler, 1993).

Besner and Care (2003) designed the Task Set paradigm to examine the role of *intention.* However, the paradigm does not necessarily distinguish between intention and attention. Given that target processing is delayed (i.e., additive effects are observed between SOA and a manipulated factor) if processing must wait until a task set in is place, *or* if processing depends on some limited capacity central resource shared with cue processing (much like in the PRP paradigm). In Experiments 1 and 4 I observed no interaction between nonword complexity and SOA. Reynolds and Besner (2006) reported this same result, but in the context of the PRP paradigm. This is consistent with the hypothesis that the delay when reading nonwords in the task set paradigm is due to the need for some limited capacity resource that cue processing also draws on. It is important to note, that even if this hypothesis is correct, a task set must be in place before nonword processing begins, otherwise subjects would need to process both tasks on every trial. As discussed previously, there is no evidence for this possibility in any of the experimental results. Thus the order of events on reading aloud trials must be (1) cue decoding (2) task set implementation and (3) sublexical processing. Thus it might be argued that sublexical processing uses both attention *and* intention. See Table 6 for a summary of the results that have been observed when reading aloud in the context of the PRP paradigm and the Task set paradigm.

When an underadditive interaction is observed between SOA and a manipulated factor in the Task Set paradigm then target processing does not need a task set to be in place and is not limited by capacity, and so can be said to need neither attention nor intention. This pattern was

	Task Set Paradigm	**PRP Paradigm**
Sublexical Processing	**Additive Effects** Nonword Complexity + SOA *(Experiments 1 and 4 here)*	**Additive Effects** Nonword Complexity + SOA *(Experiments 2 & 3 in Reynolds & Besner, 2006)*
Lexical Processing	**Under-Additive Interaction** Word Frequency × SOA *(Experiments 2,3 & 5 here)*	**Under-Additive Interaction** Word Frequency × SOA *(Experiments 1 - 3 in Cleland et al., 2006)*
	Additive Word Frequency + SOA when words are mixed with nonwords *(Experiment 4 here)*	**Under-Additive Interaction** Repetion Priming × SOA *(Experiment 1 in Reynolds & Besner, 2006)*
Semantic Processing	**Additive** Word Frequency + SOA in antonym generation *(Experiment 5 here)*	**Under-Additive Interaction** Valence congruency × SOA *(Experiment 1 in Fischer & Schubert, 2008)*

Table 6. Summary of results observed in the context of the Task Set Paradigm and in the PRP paradigm when processing by recourse to sublexical, lexical or semantic routines, as indexed by the indicated factors. The term “underadditive” is used to represent an interaction in which the effect of the manipulated factor is smaller at the shorter SOA than at the longer SOA.

indeed observed in Experiments 2 and 3, such that the effect of word frequency was underadditive with SOA. This is also consistent with results from the PRP paradigm. Cleland, Gaskell, Quinlan and Tamminen (2006) reported an underadditive interaction between word frequency and SOA in the context of the PRP paradigm There have also been reports that this interaction depends on reading skill (i.e., more skilled readers show an interaction between these factors, Ruthruff, Allen Lien & Grabbe, 2008; but see McCann, Remmington & Van Selst, 2000). Additionally, Reynolds and Besner (2006) and O'Malley et al. (2008) reported an underadditive interaction between long-lag repetition priming of words and SOA. Thus overall it appears that at least some lexical processing does not require central attention.

However in Experiment 4 of the present series, I observed no interaction between word frequency and SOA. It is difficult to see why this additivity should be attributed to the need for some limited capacity resource. As well, much of the evidence from the PRP literature also suggests that lexical processing does not need central capacity.[7] In contrast one's goals and intentions are often changing, and indeed *should* change across contexts. It is simpler then to imagine how a task set might be required in one context but not in another. In short, it appears that lexical processing does not require central *attention*, but can be affected by *intention*.

Does semantic processing require attention or intention? There has been almost no research examining the role of central attention on semantic processing when reading. An

---

[7] It remains to be seen if additivity would be observed between a lexical factor and SOA in the PRP paradigm if nonwords were intermixed in the stimulus set (as in Experiment 4). O'Malley, Reynolds, Stolz and Besner (2008) reported a PRP experiment which used long-lag repetition priming of words and pseudohomophones (nonwords that sound like real words when said aloud, e.g., brane). However, words and pseudohomophones were blocked, and so this does not provide insight into whether mixing stimulus types, in a single block, would have any effect.

exception is Fischer and Schubert (2008) who had subjects evaluate the valence (positive/negative) of a target word as Task 2 in the PRP paradigm (Task 1 was tone identification). The target was flanked above and below by two distracter words that were either congruent or incongruent in valence. The effect of congruency (RTs are faster when the distracter items are congruent vs. incongruent) interacted with SOA such that there was a smaller effect of congruency when there was temporal overlap between Task 1 and Task 2. This suggests that at least some semantic processing can occur while central attention is occupied by another task, though it should be noted that this underadditivity was by no means complete (being reduced by about 50% at the short SOA). This task is quite different that the one used in Experiment 5 (antonym generation), and so it is an empirical question as to whether, when using the same task as Fischer and Schubert, an underadditive interaction between congruency and SOA would be observed in the task set paradigm. Regardless, some form of semantic processing is delayed by either the need for attention or intention in the context of the Task Set paradigm (as evidenced by Experiment 5). Further research is needed in order to gain a clearer understanding of how semantic processing is affected by both attention and intention.

Future Directions

The results reported here provide a foundation for the examination of the role of intention when reading aloud. There are still several issues that remain to be examined. In the present work the task cue was always a tone, which is relatively easy to process and does not make any demands on lexical processing per se. In fact, that is why tones are often used in the PRP paradigm and the Task Set paradigm; any interference or delay in processing the word cannot be attributed to an overlap in the processing architecture used for tone processing and word processing. That said it is certainly of interest to know how the nature of the cue affects task processing. Using cues that are harder to process (e.g., a picture) or that require linguistic processing (e.g., a spoken word) would shed some light on this issue.

Another interesting line of research is to examine whether the secondary task affects processing of the primary task. In all the experiments reported here, the primary task was reading aloud, and with the exception of Experiment 5, the second task was case decision. Changing the nature of the second task (i.e., using a lexical decision task) may shed some additional insight into how context affects the role of intention. Finally, all the conclusions drawn here rely on RT and accuracy measures; using an electrophysiological measure (e.g., ERP and/or MEG) would provide more detailed information about the time course of these processes.

Experiment 4, in which words and nonwords are intermixed, serves to highlight the importance of context when reading aloud. At present, word recognition researchers pay lip service to the idea that context is important, but in practice typically ignore the issue (e.g., computational models typically do not, to date, attempt to model context effects (but see Reynolds & Besner, 2005; 2008; 2011; Reynolds, Besner & Coltheart, 2011; or when they do address the issue they fail; see Ziegler, Perry & Zorzi, 2009 vs. Besner & O'Malley, 2009). The present results are unambiguous: a radical change in how we think about the process of reading aloud, and the current way of modelling this process, is needed. Developing and implementing a theory of context certainly won't be easy, but if we wish to understand how print is translated into sound and meaning then such an effort is necessary.

Conclusions

As argued by Finkbeiner and Forster (2007), to undermine the claim that reading aloud proceeds autonomously one must demonstrate that the processes involved in the domain-specific stage of processing (here, lexical, sublexical and semantic) are affected by higher-level cognitive systems. The results presented here provide such evidence. When reading nonwords aloud, factors that directly affect sublexical processing (letter length and complexity), are additive with SOA, suggesting that sublexical processing does not begin until the task cue has been processed. Further a factor that indexes lexical processing (word frequency), is also additive with SOA provided that nonwords and words are intermixed in the experiment, suggesting that in this context lexical processing also does not begin until cue processing has been completed. Finally, some aspect of semantic processing, as required by the antonym generation task, does not begin until a task set is in place. These results, considered alongside results obtained in the PRP paradigm suggest that sublexical processing uses both attention and intention, and lexical and semantic processing do not require attention, but can be affected by intention. Overall, reading aloud is a highly context dependent process and is not automatic in the specific sense that processing begins as soon as a word is presented, and this process cannot be interfered with. These data provide some insight into the role of intention when reading aloud, and serve as a basis for further research on the role of context when reading aloud.

References

Ansari, I., & Besner, D. (2005). A role for set when naming arabic numerals: How intentionality limits (putatively automatic) performance. *Psychonomic Bulletin & Review, 12,* 1076-1081.

Arrington, C.M., Altmann, E.M. & Carr, T.H. (2003). Tasks of a feather flock together: Similarity effects in task switching. *Memory & Cognition, 31,* 781-789.

Baayen, R. H. (2009). *Analyzing linguistic data: A practical introduction to statistics.* Cambridge, MA: Cambridge University Press.

Balota, D.A. & Abrams, R.A (1995). Mental chronometry: Beyond onset latencies in the lexical decision task. *Journal of Experimental Psychology: Learning Memory and Cognition, 21,* 1289-1302.

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

Besner, D. (2001). The myth of ballistic processing: Evidence from Stroop's paradigm. *Psychonomic Bulletin & Review, 8*, 324-330.

Besner, D., & Care, S. (2003). A paradigm for exploring what the mind does while deciding what it should do. *Canadian Journal of Experimental Psychology, 57,* 311-320.

Besner, D., Moroz, S., & O'Malley, S. (2011). On the strength of connections between localist mental modules as a source of frequency-of-occurrence effects. *Psychological Science, 22,* 393-398.

Besner, D., & O'Malley, S. (2009). Additivity of factor effects in reading tasks is still a challenge for computational models: Reply to Ziegler, Perry, and Zorzi (2009). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 312-316.

Besner, D., O'Malley, S., & Robidoux, S. (2010). On the joint effects of stimulus quality, regularity, and lexicality when reading aloud: New challenges. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(3), 750-764.

Besner, D., Reynolds, M., & O'Malley, S. (2009). When underadditivity of factor effects in the psychological refractory period paradigm implies a bottleneck: Evidence from psycholinguistics. *The Quarterly Journal of Experimental Psychology, 62*, 2222-2234.

Besner, D., Risko, E.F., & Sklair, N. (2005). Spatial attention as a necessary preliminary to early processes in reading. *Canadian Journal of Experimental Psychology, 59*, 99-108.

Besner, D., & Roberts, M.A. (2003). Reading nonwords aloud: Results requiring change in the dual route cascaded model. *Psychonomic Bulletin & Review, 10*, 398-404.

Besner, D., & Smith, M. C. (1992). Models of visual word recognition: When obscuring the stimulus yields a clearer view. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 468-482.

Blais, C., O'Malley, S., & Besner, D. (2011). On the joint effects of repetition and stimulus quality in lexical decision: Looking to the past for a new way forward. Quarterly Journal of Experimental Psychology. *in press*.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813-840.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in Stroop color naming: Is word recognition automatic? *Journal of Experimental Psychology:General, 131*, 22, 220- 240.

Brysbaert, M. (2001). Prelexical phonological coding of visual words in Dutch: Automatic after all. *Memory & Cognition, 29*, 765-773.

Coltheart, M. Dual-route theories of reading aloud. (2011). In Adelman, J. (Ed). *Visual Word Recognition*. Hove, UK: Psychology Press.

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

Cleland, A.A., Gaskell, M.G., Quinlan, P.T., & Tamminen, J. (2006). Frequency effects in spoken and visual word recognition: Evidence from dual-task methodologies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 104-119.

Evans, G., Lambon, R.M., & Woollams, A. (2011) What's in a word? A parametric study of semantic influences on visual word recognition. *Psychonomic Bulletin & Review,* under review.

Finkbeiner, M., & Forster, K.I. (2007). Attention, intention and domain-specific processing. *Trends in Cognitive Sciences, 12,* 59-64.

Fischer, R., & Schubert, T. (2008). Valence processing bypassing the response selection bottleneck? evidence from the psychological refractory period paradigm. *Experimental Psychology*, *55*, 203-211.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123,* 71–99.

Grainger, J., Diependaele, K., Spinelli, E., Ferrand, L., & Farioli, F. (2003). Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29,* 1256 –1269.

Grondin, R., Lupker, S. J., & McRae, K. (2009). Shared features dominate semantic richness effects for concrete concepts. *Journal of Memory and Language*, *60,* 1-19.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108*, 356-388

Huey, E. B. (1908). *The psychology and pedagogy of reading*. Oxford, England: Macmillan.

Hurks, P M., Hendriksen, J.G.M., Vles, J.S.H., Kalff, A.C., Feron, F.J.M., Kroes, M., van Zeben, T.M., Steyaert, J., Jolles, J. et al. (2004). Verbal fluency over time as a measure of automatic and controlled processing in children with ADHD. *Brain and Cognition, 55*, 35-544.

Johnston, M. B., & Castles, A. (2003). Dissociating automatic orthographic and phonological codes in lexical access and lexical acquisition. In. S. Kinoshita & S. Lupker (Eds.), *Masked priming: The state of the art* (pp. 193–222). Hove, England: Psychology Press.

Kahan, T.A., Hengen, K.B. & Mathis, K.M. (2011). An examination of orthographic and phonological processing using the task-choice procedure. *Language and Cognitive Processes, 26,* 171-190.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language, 48,* 207-232.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Lachter, J.J., Forster, K.I., & Ruthruff, E. (2004). Forty-five years after Broadbent (1958): Still no identification without attention. *Psychological Review, 111*(4), 880 -913.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492-527.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments & Computers, 28*, 203-208.

MacLeod, C. M. (1991). Half a century of research on the stroop effect: An integrative review. *Psychological Bulletin, 109,* 163-203.

Marcel, A .J. (1983). Conscious and unconscious perception: Experiments on visual masking and word recognition. *Cognitive Psychology, 15,* 197-237.

Marí-Beffa, P., Valdés, B., Cullen, D. J. D., Catena, A., & Houghton, G. (2005). ERP analyses of task effects on semantic processing from words. *Cognitive Brain Research, 23*, 293-305.

Melara, R.D., & Mounts, J. R. (1993). Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice. *Memory & Cognition, 21*, 627-645.

McCann, R. S., & Besner, D. (1987). Reading pseudohomophones: Implications for models of pronunciation assembly and the locus of word-frequency effects in naming. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 14-24.

McCann, R.S., Remington, R.W., & Van Selst, M. (2000). A dual-task investigation of automaticity in visual word processing. *Journal of Experimental Psychology: Human Perception and Performance*, *26,* 1352-1370.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology:General, 106,* 226-254.

Neely, J.H. & Kahan, T.A. (2001). Is semantic activation automatic? A critical re evaluation. In H.L Roediger, J.S. Nairne, I. Neath, A.M. Surprenant (Eds.), *Th Nature of Remembering: Essays in Honor of Robert G. Crowder*. (pp 69-93). Washington, DC: American Psychological Association.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments & Computers, 36*, 402-407.

Nicolson, R. I., & Fawcett, A. J. (1990). Automaticity: A new framework for dyslexia research? *Cognition, 35*, 159-182.

O'Malley, S., & Besner, D. (2008). Reading aloud: Qualitative differences in the relation between stimulus quality and word frequency as a function of context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(6), 1400-1411.

O'Malley, S., Reynolds, M. G., Stolz, J. A., & Besner, D. (2008). Reading aloud: Spelling-sound translation uses central attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 422-429.

Pashler, H. (1993). *Dual-task interference and elementary mental mechanisms*. Cambridge, MA, US: The MIT Press.

Paulitzki, J.R., Risko, E.F., O’Malley, S., Stolz, J.A., & Besner, D. (2009). On the role of set when reading aloud: A dissociation between prelexical and lexical processing. *Consciousness and Cognition, 18,* 135-144.

Perry, C., Ziegler, J.C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review, 114,* 273-315.

Plaut, D.C. (2005). *Connectionist approaches to reading*. Malden: Blackwell Publishing.

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*(1), 56-115.

Posner, M.I., & Snyder, C.R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information Processing and Cognition: The Loyola Symposium*. (pp. 55 85). Hillsdale, H.J.: Erlbaum.

R Development Core Team. (2004). *R: A language and environment for statistical computing.*Vienna: R Foundation for Statistical Computing.

Rastle, K. & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin & Review, 5,* 277-282.

Rastle, K., Havelka, J., Wydell, T.N., Coltheart, M., & Besner, D. (2009). The cross script length effect: Further evidence challenging PDP models of reading aloud. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 238-246.

Reimer, J. F., Lorsbach, T. C., & Bleakney, D. M. (2008). Automatic semantic feedback during visual word recognition. *Memory & Cognition, 36*, 641-658.

Reynolds, M.G., & Besner, D. (2005). Contextual control over lexical and sublexical routines when reading English aloud. *Psychonomic Bulletin & Review, 12*(1), 113-118.

Reynolds, M.G., & Besner, D. (2006). Reading aloud is not automatic: Processing capacity is required to generate a phonological code from print. *Journal of Experimental Psychology; Human Perception & Performance, 32*, 1303-1323.

Reynolds, M., & Besner, D. (2008). Contextual effects on reading aloud: Evidence for pathway control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 50-64.

Reynolds, M.G., & Besner, D. (2011). There goes the neighbourhood: Contextual control over the breadth of lexical activation when reading aloud. *The Quarterly Journal of Experimental Psychology* (in press).

Reynolds, M.G., Besner, D., & Coltheart, M. (2011). Reading aloud: New evidence for contextual control over the breadth of lexical activation. *Memory and Cognition, 39,* 1332 – 1347.

Risko, E.F., Stolz, J.A., & Besner, D. (2011). Basic processes in reading: On the relation between spatial attention and familiarity. *Language and Cognitive Processes, 26*, 47-62.

Roberts, M. A., Rastle, K., Coltheart, M., & Besner, D. (2003). When parallel processing in visual word recognition is not enough: New evidence from naming. *Psychonomic Bulletin & Review, 10*(2), 405-414.

Ruthruff, E., Allen, P.A., Lien, M. & Grabbe, J. (2008). Visual word recognition without central attention: Evidence for greater automaticity with greater reading ability. *Psychonomic Bulletin & Review, 15*, 337-343.

Sabri, M., Melara, R. D., & Algom, D. (2001). A confluence of contexts: Asymmetric versus global failures of selective attention to stroop dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 515-537. Seidenberg, M.S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14*, 238-242.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, Amsterdam, 30*, 276-315.

Strain, E., Patterson, K., & Seidenberg, M.S. (1995). Semantic effects in single-word naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1140-1154.

Tombu, M., & Jolicœur, P. (2002). All-or-none bottleneck versus capacity sharing accounts of the psychological refractory period phenomenon. *Psychological Research, 66,* 274-286.

Van Orden, G. C., Pennington, B. F., & Stone, G. O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97*, 488- 522.

Vincent, S. B. (1912). The function of the vibrissae in the behavior of the white rat. *Animal Behavior Monographs, 1, 5*, 84-84.

Waechter, S., Besner, D., & Stolz, J. A. (2011). Basic processes in reading: Spatial attention as a necessary preliminary to orthographic and semantic processing. *Visual Cognition*, *19*, 171-202.

Weekes, B.S. (1997). Differential effects of number of letters on word and nonword naming latency. *The Quarterly Journal of Experimental Psychology, 50,* 439-456.

Wilson, L. B., Tregellas, J. R., Slason, E., Pasko, B. E., & Rojas, D. C. (2011). Implicit phonological priming during visual word recognition. *NeuroImage, 55*, 724-731.

Woollams, A.M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 878-890

Xu, B., & Perfetti, C. A. (1999). Nonstrategic subjective threshold effects in phonemic masking. *Memory & Cognition*, *27*(1), 26-36.

Yap, M. J., & Balota, D. A. (2007). Additive and interactive effects on response time distributions in visual word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 274-296.

Yap, M.J., Balota, D.A., Tse, C.S. & Besner, D. (2008). On the additive effects of stimulus quality and word frequency in lexical decision: Evidence for opposing interactive influences revealed by RT distributional analyses. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34,* 495-513.

Ziegler, J. C., Perry, C., & Zorzi, M. (2009). Additive and interactive effects of stimulus degradation: No challenge for CDP+. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(1), 306-311.

Appendix A: Nonwords used in Experiments 1 and 4

Short/Basic Nonwords

- blaf
- blem
- blif
- blun
- brol
- brup
- clek
- clig
- clil
- clis
- crel
- crem
- cren
- cril
- crut
- dreb
- drec
- drif
- dwab
- dwak
- dwep
- dwiz
- falp
- fesk
- flif
- fliz
- frid
- frub
- frup
- gelk
- glak
- glif
- glof
- glyp
- grud
- grus
- jalc
- jalp
- jelm
- julb
- kalp
- kanc
- kesk
- klel
- klen
- kred
- krum
- kulf
- lalp
- nalp
- nusp
- plib
- plic
- plif
- plil
- plis
- pliv
- pliz
- plym
- prub
- prud
- prul
- prun
- relk
- scib
- scig
- scuk
- skeb
- skol
- skos
- slel
- slez
- smeb
- smet
- smib
- smif
- smik
- snef
- snel
- snet
- snez
- snis
- spad
- spiv
- spom
- spuk
- srec
- srep
- staz
- stec
- stet
- stob
- stum
- trel
- trif
- trub
- tulb
- velk
- visk
- wumf
- yalc
- zalp
- zelk

Long/Complex Nonwords

- blarch
- bleece
- bloide
- blooch
- blynch
- brauce
- browth
- claete
- cleace
- clouch
- clouse
- craith
- creeph
- creeze
- criege
- croche
- drelch
- drodge
- drowth
- dwache
- dwirch
- dwitch
- dwudge
- fautch
- feague
- fladge
- flenge
- fralph
- freich
- frouse
- gautch
- gladge
- gleece
- gleigh
- glough
- grodge
- groose
- grouge
- jautch
- jeathe
- jorgue
- jourth
- kauche
- kleigh
- klough
- krieve
- kroupe
- kusque
- lourth
- nounge
- nourge
- plaesh
- plault
- plawce
- plawgh
- plawle
- pleege
- pleeph
- ploice
- praele
- preece
- preige
- priesh
- rauche
- scawce
- sceeve
- scight
- skedge
- skeuth
- skinch
- sleege
- slouse
- smaefe
- smaice
- smawsh
- smeave
- smeigh
- snaeph
- snauge
- sneave
- snooge
- snooth
- speuce
- sprate
- sprine
- sprune
- sprush
- sreeve
- sreeze
- stribe
- strine
- strofe
- stroge
- strone
- toathe
- trouge
- trouph
- trouse
- vautch
- vounge
- worgue
- yauche
- zaitch
- zeathe

Appendix B: Low Frequency words used in Experiments 2 to 5

Low Frequency Words (Antonym)

absent (present)
abundant (scarce)
acquit (convict)
agony (extasy)
aloof (engage)
amplify (reduce)
anarchy (government)
antidote (poison)
ascend (descend)
attract (repulse)
aunt (uncle)
autumn (spring)
awake (asleep)
backward (forward)
bend (straighten)
bitter (sweet)
bless (curse)
blunt (sharp)
blur (clear)
bold (timid)
brave (scared)
bride (groom)
calm (nervous)
casual (formal)
catcher (pitcher)
cheerful (sad)
combine (separate)
comfort (discomfort)
conceal (reveal)
convex (concave)
cooked (raw)
cruel (kind)
defend (accuse)
deposit (withdraw)
despair (hope)
dim (bright)
divorce (marriage)
dull (sharp)
dusk (dawn)
elated (sad)
elder (younger)
elude (seek)
emerge (submerge)
eternal (mortal)
exhale (inhale)
exotic (ordinary)
expand (contract)
export (import)
fake (real)
fancy (plain)
fickle (loyal)
float (sink)
forbid (allow)
forgive (blame)
frown (smile)
genuine (fake)
goodbye (hello)
graceful (clumsy)
guilty (innocent)
hasten (dawdle)
heaven (hell)
height (depth)
hero (coward)
hide (seek)
hollow (solid)
hunger (thirst)
implicit (explicit)
imprison (free)
indoor (outdoor)
inferior (superior)
inflate (deflate)
inhabit (uninhabit)
inhale (exhale)
interior (exterior)
isolate (include)
jerky (smooth)
junior (senior)
lazy (active)
lend (borrow)
lofty (lowly)
loyal (disloyal)
luxury (squalor)
mature (immature)
mediocre (great)
merry (mirthless)
miser (spendthrift)
misses (hits)
multiply (divide)
naive (worldly)
narrow (wide)
neat (messy)
niece (nephew)
noisy (quiet)
noon (midnight)
obey (command)
opaque (transparent)
optimist (pessimist)
ornate (plain)
outer (inner)
owe (pay)
perish (survive)
polite (rude)
precious (worthless)
prey (preditor)
pride (humility)
puny (stout)
rapid (slow)
reap (sow)
reckless (cautious)
reject (accept)
relax (tense)
retreat (advance)
reward (punish)
rigid (soft)
rise (fall)
rival (friend)
rural (urban)
scarce (plenty)
scarce (plentiful)
scatter (collect)
seldom (often)
shallow (deep)
shiny (dull)
sink (float)
skinny (fat)
sleazy (smooth)
sober (drunk)
sorrow (joy)
squander (save)
suave (clumsy)
subtract (add)
sudden (gradual)
sunny (cloudy)
tall (short)
tame (wild)
tragedy (comedy)
vacant (occupied)
victory (defeat)
villain (hero)
virtue (vice)
vowel (consonant)
wake (sleep)
whisper (yell)

Appendix C: High Frequency words used in Experiments 2 to 5

High Frequency Words (Antonym)

above (below)
accurate (inaccurate)
adult (child)
advantage (disadvantage)
after (before)
against (for)
agree (disagree)
ahead (behind)
alive (dead)
all (none)
always (never)
ancient (young)
answer (question)
attack (defense)
attention (inattention)
beautiful (ugly)
before (after)
beginning (ending)
behind (infront)
better (worse)
black (white)
blame (praise)
bottom (top)
break (fix)
brother (sister)
build (destroy)
careful (careless)
cause (effect)
cheap (expensive)
city (country)
clean (dirty)
cold (hot)
common (rare)
complete (incomplete)
cool (warm)
correct (wrong)
dark (light)
daughter (daughter)
desire (spurn)
different (same)
difficult (easy)
doctor (patient)
driver (passenger)
early (late)
employer (employee)
end (begin)
escape (capture)
even (odd)
external (internal)
famous (unknown)
fast (slow)
father (mother)
find (lose)
first (last)
floor (ceiling)
follow (lead)
found (lost)
freedom (captivity)
friend (enemy)
full (empty)
funny (serious)
future (past)
girl (boy)
give (take)
good (bad)
guest (host)
happy (sad)
hard (soft)
heavy (light)
here (there)
high (low)
include (exclude)
increase (decrease)
input (output)
inside (outside)
king (queen)
knowledge (ignorance)
land (sea)
least (most)
left (right)
less (more)
life (death)
long (short)
loose (tight)
loss (gain)
lots (little)
love (hate)
master (servant)
maximum (minimum)
miss (mister)
near (far)
never (always)
new (used)
nice (mean)
north (south)
obvious (unclear)
off (on)
open (close)
opinion (fact)
original (copy)
over (under)
parent (child)
pass (fail)
peace (war)
permanent (temporary)
please (displease)
pleasure (pain)
plus (minus)
poor (rich)
possible (impossible)
pretty (ugly)
private (public)
pull (push)
qualified (unqualified)
question (answer)
quick (slow)
random (ordered)
real (fake)
receive (give)
reduce (increase)
remember (forget)
rich (poor)
safe (danger)
separate (join)
serious (trivial)
simple (complex)
solid (liquid)
stable (unstable)
start (finish)
stop (go)
strong (weak)
talk (listen)
teach (learn)
truth (lie)
valley (hill)
waste (use)
win (lose)
within (without)
work (play)
wrong (right)
yes (no)
yesterday (tomorrow)