**Variables**

The variables used in this study can be grouped into one of four broad categories: lexical measures, associations, semantics, and thematics. Lexical measures refer to any variables which describe the physical make up of a word (i.e., word length, part of speech) or its place relative to other words (i.e., frequency or neighborhood connections). Next, we are interested in examining the effects of prime-target relatedness on priming. As “relatedness” is a broad term that can describe several types of relationships between concepts, we have divided these variables into three categories. First, we define associative measures as those variables which capture information pertaining to the use of words together in context. (i.e., *Bird – Song)*. Next, semantic measures capture information pertaining to a word’s meaning, often in terms of shared features (i.e., *Bird – Pigeon*). Finally, thematic measures refer to variables which capture information about word relationships centered around the links between concepts and an overarching scenario or theme. (i.e., *Bird – Nest)*. These measures are described in further detail below.

**Lexical Measures**

The lexical properties chosen for this study were frequency, length, and orthographic and phonographic neighborhood sizes. Word frequency has been shown to affect semantic priming. Becker (1979) showed larger priming effects for low-frequency words than high frequency words. Word frequency information was taken from the SUBTLEX project (Brysbaert & New, 2009). Next, word length was calculated as the number of characters comprising a word. Previous research suggests that longer words require more time to be processed (Kuperman et al., 2012) and that [SECOND EXAMPLE HERE]. Finally, orthographic and phonographic neighborhood information for primes and targets was taken from the Semantic Priming Project (Hutchison et al., 2013). Furthermore, previous work by Hutchison et al. (2008) suggests that semantic priming is greatest when related primes are short in length and share few orthographic neighbors.

**Relatedness Measures**

Three measures of word association were examined in this study. First, forward strength (FSG) values were taken from the Small World of Words project (SWOW, De Deyne et al., 2018; De Deyne et al., 2013) to be used as a measure of associative overlap between words. The SWOW norms consist of association data for over 12,000 English cue words collected from approximately 90,000 respondents (De Deyne et al., 2018). This set of norms differs from more traditionally used sets of association data (i.e., the University of South Florida Free Association Norms – Nelson et al., 2004) as they are generated through a multiple response free association task, rather than through single responses. Each SWOW cue item receives three responses from each participant. This multiple response procedure generates a more diverse response set, which allows for better predictions of item relatedness and lexical access than traditional single-response procedures (De Deyne et al., 2013).

In addition to associative overlap, we also examined the effects of two measures of associative network size. FSG set size is defined as number of times that a word is the cue item, and FAN set size is defined as the number of times that a word is the target item. Previous work by Hutchison (2003) suggests that priming based on associations can occur in the absence of feature of overlap. Additionally, Heyman et al. (2017) found that FSG led to different levels of priming with changes to working memory load.

Next, measures of semantic relatedness were investigated. Cosine values for roots and affixes were calculated from the Buchanan et al. (2013) semantic word norms (i.e., the root word *abandon* may take several forms, such as *abandoned*, *abandoning*, etc.). Information for feature set sizes (defined as the number of related features a word has) and cosine set sizes (defined as the number of times a word is the cue item) was also collected.

The final group of overlap variables are those which capture aspects of thematic relationships between pairs (i.e., contextual similarity). While thematic word relations are a facet of semantic memory, research by Maki and Buchanan (2008) suggests that themes comprise a separate area of linguistic processing. Additionally, previous research by Jones et al. (2006) has found distributional models of semantic memory to predictive of priming effects. Such models of semantic memory (i.e., LSA, HAL, BEAGLE) provide a means of measuring thematic overlap, as these models derive similarity of meaning from co-occurrences in text rather than purely in terms of shared features. For the present study, we examined the effects of thematic overlap by examining norms derived from three distributional models.

First, we examined LSA (Landauer and Dumais, 1997; Landauer et al., 1998), as Hutchison et al. 2008 found LSA strength to be predictive of overall priming. Next, we included values derived from BEAGLE (Jones & Mewhort, 2007), and a measure of semantic distance derived from Mandera’s Continuous Bag of Words model (CBOW, Mandera et al., 2017). BEAGLE is best described as a model that creates a semantic space representation of meaning based upon word order and statistical redundancies inherent in language. Text is processed one sentence at a time, allowing the model to learn order information pertaining to each word and the context in which it is used (Jones & Mewhort, 2007). Similarly, CBOW models work by using context words to predict the current word (Mandera 2017).

Maybe a final paragraph on Hutchison stuff about “pure” pairs being hard to separate out?