Assessing automatic processing of hypernymic relations in first language speakers and advanced second language learners

A semantic priming approach

Scott Crossley
Georgia State University

This study investigates the depth of lexical knowledge in first language (L1) speakers and second language (L2) learners in reference to hierarchical word knowledge. Eighty-eight participants took part in a lexical decision task that assessed their speed and accuracy in recognizing words and nonwords. Prime and target pairs in the lexical decision task were related words (hyponym to hypernym and hypernym to hyponym), unrelated words, or word to nonwords. The findings indicate bidirectional priming in L1 participants such that associated pairs (hyponym to hypernym and hypernym to hyponym) were processed faster than unrelated words. For L2 participants, unidirectional priming effects were reported for the hyponym to hypernym condition only. These findings provide evidence that hierarchical lexical networks characterize L1 lexicons but not L2 networks. Such findings provide important information about the organizational properties of L1 and L2 lexicons.

Keywords: semantic priming, hypernymy, lexical networks, second language learning, lexical competence, depth of lexical knowledge

Assessing the lexical competence of English speakers is relatively straightforward for many features of word knowledge (e.g., evaluating breadth of knowledge features such as word frequency and lexical diversity); however, some aspects of lexical competence can only be assessed indirectly (e.g., depth of knowledge features such as collocational knowledge and hierarchical relations between words). These areas of lexical knowledge are generally the result of incidental learning that occurs when speakers focus their attention on word meaning and not word form.
Assessing automatic processing of hypernymic relations

(Nation, 2001; Schmitt, 1998; VanPatten & Williams, 2007). For first language (L1) speakers of English, word knowledge is, by and large, acquired implicitly and fluently. For second language (L2) speakers of English, word knowledge develops at a slower rate and involves both implicit and explicit learning. Fluency is generally not expected for L2 learners, but lexical competence at a variety of levels, to include depth of knowledge, is attainable.

One approach to assessing a speakers’ depth of lexical knowledge is through corpus and computation analyses of spoken or written data (e.g., Crossley, Salsbury, & McNamara, 2009, 2010a). A second approach is through direct assessment of lexical knowledge (i.e., tests of productive and passive word knowledge; Meara, 2005; Schmitt, 1998). A final approach, and potentially the most promising, is through the use of semantic priming experiments that assess implicit links between related words. Semantic priming experiments investigate the tendency for L1 and L2 users to process a word more accurately and quickly immediately after they have been exposed to a related word. For instance, users will identify, as a word, cat more quickly if they have recently been exposed to the word dog as compared to if they have been exposed to an unrelated word such as pen (McDonough & Trofimovich, 2009). Semantic priming experiments are used to assess how previous language exposure influences lexical processing through an analysis of how words prime lexical chains (i.e., dog and cat). Many researchers argue that semantic priming reflects the fundamental properties of lexical organization and the manner in which words are stored and retrieved (Neely, 1991; Ratcliff & McKoon, 1988). When used to assess depth of lexical knowledge, priming tasks also have an ecological validity because they depend on implicit processes that tap into implicit memory (McDonough & Trofimovich, 2009). Previously, priming experiments have been used to assess the lexical competence of L1 speakers and bilinguals in a variety of tasks related to depth of lexical knowledge (Devitto & Burgess, 2004; Frenck-Mestre & Prince, 1997; McNamara, 2005; Meyer & Schvaneveldt, 1971); however, their use in L2 environments has been rare.

The purpose of this study is to use semantic priming techniques to assess the strength of hypernymic relations (i.e., links between superordinate and subordinate terms in hierarchies such as weapon → gun → rifle) in the lexicons of L1 and L2 speakers. Such techniques allow for the assessment of the lexical competence of L1 and L2 speakers in terms of their depth of lexical knowledge. A semantic priming task was developed and tested on both an L1 and an L2 population. The priming task affords the opportunity to assess the lexical knowledge of L1 and L2 speakers in reference to priming effects for both hypernyms (superordinate terms) and hyponyms (subordinate terms), which, to my knowledge, have not been examined. Additionally, the priming task affords a comparison of priming effects between the baseline L1 population and the L2 population. Such a
comparison provides an opportunity to understand potential differences in the organization of L1 and L2 lexicons.

**Lexical Competence**

Assessing lexical competence is important because it is an essential element of academic success (Daller, van Hout, & Treffers-Daller, 2003). Lexical knowledge forms the basis for more advanced language skills, including syntax (Ellis, N., & Ferreira-Junior, 2009; Wulff, Ellis, Romer, Bardovi-Harlig, & LeBlanc, 2009) and reading comprehension (Hu & Nation, 2000). Pragmatically, the lexicon is also crucial for communicative success with misinterpretations of lexical items identified as key elements in communication errors (Crossley, Salsbury, & McNamara, 2010b; Ellis, R., 1995; Ellis, R., Tanaka, & Yamazaki, 1994; de la Fuente, 2002). From a theoretical perspective, understanding L2 lexical acquisition in relation to its deeper, cognitive functions can lead to increased awareness of how language users process and produce language (Crossley et al., 2009, 2010a).

A key element of lexical competence is depth of lexical knowledge (i.e., the degree of organization of known words; Meara, 1996, 2005; Qian, 1999; Read, 1998; Wesche & Paribakht, 1996). Unlike breadth of lexical knowledge, depth of lexical knowledge is not based on the number or variety of words a learner produces, but on constraints at the phonemic, morphemic, and syntactic level and at deeper levels related to word associations (e.g. semantic co-referentiality, hypernymy, polysemy, and collocation knowledge; Qian & Schedl, 2004). Word association features, which are of primary interest for this study, are often integrated under the term “lexical network,” which serves as a convenient metaphor to describe the manner in which lexical features combine to form complex association models that act categorically to form entire lexicons (Haastrup & Henriksen, 2000; Huckin & Coady, 1999). From an acquisitional perspective, as language learners develop lexical competence, they build lexical networks that are strengthened by differentiating relations between words and within words (Haastrup & Henriksen).

**Lexical Networks and Semantic Priming**

A key component in semantic priming theory is the notion that lexical networks result from spreading activation. Spreading activation theory argues that activation among related semantic concepts in memory results in priming effects for associated concepts, which are represented as interconnected nodes. The more similar the concepts are, the closer they are stored to one another. The more distant the concept, the farther apart they are stored and the fewer number of links
that are shared. Within the network, the activation of a single concept will trigger the activation of all its interconnected conceptual nodes (Collins & Loftus, 1975). Findings from a variety of priming studies support spreading activation theory in that strongly associated words appear to be stored together or, at minimum, to be linked in the mind of the language user. As a result, exposure or production of a word primes the activation of related words (Neely, 1991), resulting in a priming effect in which related words are processed and identified more quickly. Alternatively, priming may not be based on associations between discrete items (i.e., individual words), but through sets of semantic features shared between that facilitate processing (McRae & Boisvert, 1998; Plaut, 1995). Thus, spreading may not be the result of associative relations between words alone (e.g., cradle-baby), but rather shared semantic features (e.g., the shared semantic features found in Zebra-Horse; Perea & Rosa, 2002).

While semantic priming is a common approach to assessing the lexical knowledge of L1 speakers (see McNamara, 2005 for an overview), it has been used less frequently in assessing the lexical knowledge of L2 learners, with most studies focusing on bilingual lexicons. Two primary research strands are found in L2 priming studies: (1) investigations to assess if similar patterns of priming exist in both the L1 and the L2; and (2) investigation into whether the priming patterns in bilinguals and L2 speakers resemble those found in monolingual speakers (McDonough & Trofimovich, 2009). In reference to the first strand, research has indicated that priming effects are stronger in speakers’ L1 than their L2, especially when proficiency differences exist between the languages (Phillips, Segalowitz, O’Brien, & Yamasaki, 2004). In reference to the second strand, research indicates that L1 speakers and advanced L2 speakers access and use some aspects of semantic information in a similar manner for strongly associated words (antonyms, synonyms, and collocations, Frenck-Mestre & Prince, 1997) but not weakly associated words (Devitto & Burgess, 2004).

The majority of within language priming studies in bilingual and L2 research have focused on semantically related words such as sugar-sweet (see McDonough & Trofimovich, 2009 for an overview) or collocations (Frenck-Mestre & Prince, 1997). A few studies have also examined synonyms and antonyms (Dong, Gui, & MacWhinney, 2005; Frenck-Mestre & Prince, 1997), but none, to my knowledge, have examined conceptual or hierarchical lexical knowledge (e.g., priming effects between superordinate and subordinate words). Such relations between words have been hypothesized to be the single most important organizational system for lexical relations (Miller & Teibel, 1991).
Hypernymic Relations

Hypernymic relations are hierarchical associations between hypernyms (superordinate words) and hyponyms (subordinate words). A hypernym is less specific than a related word (animal compared to dog) and a hyponym is more specific than a related word (dog as compared to animal). A good example of hypernymy is the relationship between car and vehicle in which car is the hyponym of the hypernym vehicle because car has a narrower and more specific definition than vehicle. Hypernymic relations allow for hierarchical categorizations that define how hyponyms inherit properties from their related hypernyms and allow set inclusion among category members. Hypernymy is a foundational lexical relationship that is consistent with network models that allow for the economical representation of lexical properties (Chaffin & Glass, 1990; Murphy, 2004). Such properties allow for learners to generalize terms as well as generate cognitive economy because every object is part of a conceptual category and not isolated (Murphy). A categorical relation such as hypernymy that is based on an economy of representations should afford faster processing of language as compared to sense relations such as synonymy or polysemy, which are argued to be more difficult to process (Chaffin & Glass).

Knowledge of hypernymic relations between words is associated with a speaker’s aptitude for managing academic and formal registers (Snow, 1990; Snow, Cancino, Gonzalez, & Shriberg, 1989; Snow, Cancino, De Temple, & Schley, 1991). For instance, the more informal speech is, the more hypernyms are used (i.e., I drove the car there), and the more formal speech is, the more hyponyms are used (i.e., I drove the Cadillac to the country club; Ordonez, Carlo, Snow, & McLaughlin, 2002). From a developmental perspective, hypernymic relations are more likely acquired as the learners advance cognitively (Anglin, 1993; Snow, 1988; Vygotsky, 1962), as they increase their levels of education (LeVine, 1980; Snow, 1990), as they acquire more specific lexical knowledge (Wolter, 2001), and in an L1 rather than an L2 (Meara, 1982; Söderman, 1993). Studies have demonstrated that at a very young age, children can distinguish between hypernyms and hyponyms in their L1 and that children first develop superordinate categories (Berlin, Breedlove, & Raven, 1974; Brown, 1958; Mervis & Crisafi, 1982; Murphy, 2004) followed by subordinate categories (Berlin et al., 1974).

Second language research has demonstrated that L2 learners do not have the same access to hypernymic relations as L1 speakers. For instance, Levenston and Blum (1977) found that L2 learners use more words of general than of specific meanings, causing the L2 learners to make inappropriate overgeneralizations and to not adhere to expected uses of registers and collocations. Ijaz (1986) found that L2 learners associated prepositions most often with contexts that were prototypical, but they failed to attribute similar relationships to non-prototypical
categories. More recent studies have demonstrated that L2 learners produce more words that are less specific as time is spent studying English (Crossley et al., 2009) indicating that L2 learners may first learn general words and then move toward the production of hypernyms.

Methods

The purpose of this study is to examine associative priming effects for hypernyms and hyponyms in a within-language priming context in order to investigate whether hierarchical relations between words influence how words are stored and processed in the mental lexicons of advanced L1 speakers and L2 speakers. In this study, L1 and advanced L2 speakers of English were given a lexical decision task to complete in which priming effects for both hypernyms and hyponyms were tested. A semantic priming approach was selected because it is argued to test automatic processes and to reflect fundamental properties of speakers’ mental lexicons such as how the lexicon is accessed and how words are retrieved. Semantic priming approaches also afford the examination of whether prior exposure to a related word facilitates processing (McDonough & Trofimovich, 2009). It is predicted that hypernyms and hyponyms will show significant facilitative effects for lexical decisions involving related words. This hypothesis is premised on the presumed centrality of hypernymic relations in developed lexicons (Chaffin & Glass, 1990; Miller & Teibel, 1991) and on the notion that less economical sense relations (i.e., synonymy; Chaffin & Glass, 1990) have demonstrated priming effects for both L1 and L2 speakers (Frenck-Mestre & Prince, 1997).

Participants

Eighty-four undergraduate students from Georgia State University volunteered for the experiment. Participants were assigned to two language conditions according to their self-identification as either second language (L2) speakers of English or first language (L1) speakers of English. Forty-one of the participants were L2 speakers of English and 43 of the participants were native speakers of English. All students received class credit in a freshman Psychology course for participating in the experiment.

Demographic data was self-reported by the participants. The average age for the L1 speakers was 21 years old. The average grade point average (GPA) for the L1 participants was 3.24. Thirty-seven percent of the L1 speakers in the study were male, while the other 63% were female. The average age of the L2 speakers
was 25 years old and they also reported an average GPA of 3.24. As with our L1 participants, 37% of the L2 speakers in the study were male, while the other 63% were female. The L2 speakers came from 22 different L1 backgrounds of which the three most common were Spanish (ten L1 speakers), Chinese (five L1 speakers), and Urdu (three L1 speakers). On average, the L2 participants had studied English for 12 years and began studying English at the age of 12.

All L2 speakers admitted to Georgia State university are required to score above a 430 on the verbal section of the Scholastic Aptitude Test (SAT) or a 17 on the English section of the American College Test. If the L2 speakers reside in a country other than the United States where English is not the dominant language, they must either score a 550 on the TOEFL exam or pass through six levels of the University’s Intensive English Program (generally a year-long program). Knowing these constraints, all L2 speaking participants were classified as advanced level English speakers. For those L2 participants that reported a TOEFL score, the average score for the TOEFL Internet-Based Test was 106.666 ($n = 12$). The average score for the paper-based test was 595.666 ($n = 3$).

**Materials and Study Design**

For this experiment, 60 prototypical words from English (i.e., doctor, dog, job, movie) were selected as primes along with 60 related hypernyms and 60 related hyponyms as target words. We operationalized prototypicality as a function of word length and word frequency (under the presumption that more prototypical words would be shorter and more frequent, Vergès, 1992). The hypernyms and hyponyms for each prototypical word were selected from the WordNet database (Fellbaum, 1998; Miller, Beckwith, Fellbaum, Gross, & Miller, 1990), which provides lists of hypernyms and hyponyms for over 100,000 English words. The principle semantic category for the majority of the 180 primes and targets was noun; however, some of the words had multiple semantic categories (i.e., the words could function as nouns and verbs such as ground, name, and drink). None of the selected words contained inflectional or derivational endings. Two experts reviewed the word pairs to ensure that phonological similarities among the pairs were minimal. Sixty nonwords were also randomly selected from ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). These nonwords were used as targets in the lexical decision task in which participants were asked to judge if a string of letters was an English word or a nonword. All nonwords contained orthographically existing onsets and bodies and thus followed the phonotactic structure of English. None of the nonwords contained inflectional or derivational endings. Table 1 contains a list of the words and nonwords used in this experiment.
### Table 1. Words and Nonwords Used in Experiment

<table>
<thead>
<tr>
<th>Prototype Prime Words</th>
<th>Hyponym to Hypernym Condition</th>
<th>Hypernym to Hyponym Condition</th>
<th>Nonwords</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEAM</td>
<td>UNIT</td>
<td>CREW</td>
<td>CRARP</td>
</tr>
<tr>
<td>CAR</td>
<td>VEHICLE</td>
<td>AMBULANCE</td>
<td>SKERK</td>
</tr>
<tr>
<td>COUNTRY</td>
<td>TERRITORY</td>
<td>CHINA</td>
<td>SWEEL</td>
</tr>
<tr>
<td>BOSS</td>
<td>LEADER</td>
<td>FOREMAN</td>
<td>NIPE</td>
</tr>
<tr>
<td>PARTY</td>
<td>AFFAIR</td>
<td>DANCE</td>
<td>KNARSE</td>
</tr>
<tr>
<td>GUN</td>
<td>WEAPON</td>
<td>RIFLE</td>
<td>CLEINT</td>
</tr>
<tr>
<td>FRIEND</td>
<td>PERSON</td>
<td>BUDDY</td>
<td>VISPE</td>
</tr>
<tr>
<td>HOSPITAL</td>
<td>STRUCTURE</td>
<td>ASYLUM</td>
<td>TWURN</td>
</tr>
<tr>
<td>MONEY</td>
<td>CURRENCY</td>
<td>DOLLAR</td>
<td>RUP</td>
</tr>
<tr>
<td>OFFICE</td>
<td>PLACE</td>
<td>ROOM</td>
<td>DEUD</td>
</tr>
<tr>
<td>TOWN</td>
<td>LOCATION</td>
<td>PARIS</td>
<td>WHUFF</td>
</tr>
<tr>
<td>DOCTOR</td>
<td>SCHOLAR</td>
<td>SURGEON</td>
<td>GLERTH</td>
</tr>
<tr>
<td>DREAM</td>
<td>FANTASY</td>
<td>NIGHTMARE</td>
<td>DWEF</td>
</tr>
<tr>
<td>YELLOW</td>
<td>COLOR</td>
<td>GOLD</td>
<td>SMOAP</td>
</tr>
<tr>
<td>BAR</td>
<td>AREA</td>
<td>LOUNGE</td>
<td>SLOUR</td>
</tr>
<tr>
<td>DECISION</td>
<td>CHOICE</td>
<td>RESOLUTION</td>
<td>PRINE</td>
</tr>
<tr>
<td>DAY</td>
<td>TIME</td>
<td>TOMORROW</td>
<td>CLALK</td>
</tr>
<tr>
<td>SEAT</td>
<td>SPACE</td>
<td>CHAIR</td>
<td>GRUCH</td>
</tr>
<tr>
<td>NEWS</td>
<td>INFO</td>
<td>REPORT</td>
<td>PAFT</td>
</tr>
<tr>
<td>WORLD</td>
<td>OBJECT</td>
<td>NATURE</td>
<td>DRERGE</td>
</tr>
<tr>
<td>WOMAN</td>
<td>FEMALE</td>
<td>WIFE</td>
<td>CLERN</td>
</tr>
<tr>
<td>FATHER</td>
<td>PARENT</td>
<td>DAD</td>
<td>SMIN</td>
</tr>
<tr>
<td>PEACE</td>
<td>ORDER</td>
<td>TRUCE</td>
<td>FRING</td>
</tr>
<tr>
<td>LOVE</td>
<td>EMOTION</td>
<td>WORSHIP</td>
<td>WROOB</td>
</tr>
<tr>
<td>PICTURE</td>
<td>ART</td>
<td>PHOTO</td>
<td>CHEUTH</td>
</tr>
<tr>
<td>HOUSE</td>
<td>HOME</td>
<td>CABIN</td>
<td>GWEIGH</td>
</tr>
<tr>
<td>GIFT</td>
<td>POSSESSION</td>
<td>PRIZE</td>
<td>ZIM</td>
</tr>
<tr>
<td>SCHOOL</td>
<td>INSTITUTION</td>
<td>ACADEMY</td>
<td>FRAWP</td>
</tr>
<tr>
<td>MOVIE</td>
<td>SHOW</td>
<td>MUSICAL</td>
<td>SPRORD</td>
</tr>
<tr>
<td>WIND</td>
<td>WEATHER</td>
<td>BREEZE</td>
<td>PHLOLL</td>
</tr>
<tr>
<td>KID</td>
<td>JUVENILE</td>
<td>ORPHAN</td>
<td>PSAWK</td>
</tr>
<tr>
<td>BAG</td>
<td>CONTAINER</td>
<td>PURSE</td>
<td>WRUMF</td>
</tr>
<tr>
<td>PLAN</td>
<td>IDEA</td>
<td>PLOT</td>
<td>KNUD</td>
</tr>
<tr>
<td>STREET</td>
<td>ROAD</td>
<td>ALLEY</td>
<td>SNIMPSE</td>
</tr>
<tr>
<td>DOG</td>
<td>ANIMAL</td>
<td>PUPPY</td>
<td>FURK</td>
</tr>
<tr>
<td>HAIR</td>
<td>COVERING</td>
<td>BEARD</td>
<td>GWERB</td>
</tr>
<tr>
<td>SONG</td>
<td>MUSIC</td>
<td>CHANT</td>
<td>SHOAF</td>
</tr>
<tr>
<td>GAME</td>
<td>CONTEST</td>
<td>FOOTBALL</td>
<td>FOWN</td>
</tr>
<tr>
<td>CLASS</td>
<td>EVENT</td>
<td>LECTURE</td>
<td>PHRALT</td>
</tr>
<tr>
<td>GROUND</td>
<td>MATERIAL</td>
<td>DIRT</td>
<td>FAIVE</td>
</tr>
</tbody>
</table>
The words were arranged into three equal groups with each group consisting of 20 prime-hypernym pairs (the hyponym to hypernym condition; e.g., team-unit), 20 prime-hyponym pairs (the hypernym-hyponym condition; e.g., team-crew), 20 hypernym-hyponym unrelated pairs (the unrelated condition; e.g., unit-Paris), and 60 words selected from the remaining primes (n = 20), hypernyms (n = 20), and hyponyms (n = 20) that were paired with the 60 nonwords (the nonword condition; e.g., unit-skerk). As a result, each of the three groups contained the same divisions of conditions and the same words, but each condition in each group contained different words (except the nonword condition which always contained the same 60 nonwords). Thus, participants only saw each word once in the experiment. This arrangement also permitted an even division between target words and nonwords such that participants could not guess, based on frequency, if a target was more likely to be a word or a nonword.

All words were controlled for word frequency and word length. Word frequency counts came from the SUBTLEXus corpus developed by Brysbaert and New (2009). The SUBTLEXus frequency counts come from a corpus of 51 million words from subtitles of American television shows and films. As a result, the
SUBTLEX corpus focuses on everyday interactions and more likely reflects the distributional properties of the natural language to which L1 and L2 speakers are exposed. The primary frequency count in SUBTLEX is contextual diversity (CD), which provides frequency counts from 8,388 different contexts and provides a means to calculate frequency without overestimating word form frequencies that result from multiple occurrences in a limited number of sources or even a single source (Adelman, Brown, & Quesada, 2006; Brysbaert & New, 2009). All words selected for this study were within the 9,000 most frequent words in the English language. Analysis of variance (ANOVA) analyses demonstrated that no significant differences in word frequency or word length existed between the selected hypernyms and hyponyms. Significant differences did exist between the word frequency scores for the prototype primes and the hypernyms ($p < .001$) and the prototype primes and hyponyms ($p < .001$). Significant differences were also reported for word length between the prototype primes ($p < .001$) and the hypernyms and the prototype primes and hyponyms ($p < .001$). Descriptive statistics for this analysis are reported in Table 2.

**Apparatus and Procedure**

Display of stimuli along with response time recordings were collected on a desktop computer using E-Prime software. Stimuli were displayed in uppercase letters in the center of the screen. The trials began with a center fixation point (+) followed by the presentation of the prime word for 250 milliseconds. The prime word was then replaced immediately by the target word, which was either a hypernym or hyponym of the prime word, an unrelated word, or a nonword. The target word remained displayed until the participants provided a response (i.e., was the target a word or a nonword). Responses were collected using a serial response box. Subjects received 10 practice trials followed by a single block of 120 trials, presented in random order. Demographic information was collected prior to the experiment using the E-Prime software. The entire experiment took about 20 minutes on average.

<table>
<thead>
<tr>
<th>Index</th>
<th>Prototype primes</th>
<th>Hypernyms</th>
<th>Hyponyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Word frequency</td>
<td>2.845 (0.415)</td>
<td>2.927 (0.475)</td>
<td>3.512 (0.206)</td>
</tr>
<tr>
<td>Word length</td>
<td>5.917 (1.619)</td>
<td>6.383 (1.814)</td>
<td>4.983 (1.501)</td>
</tr>
</tbody>
</table>
Data Analysis

A variety of statistical analyses were conducted to assess whether L1 speakers and advanced L2 learners showed priming effects for hypernyms and hyponyms in both response times (i.e., speed of judgments) and accuracy (i.e., number of correct judgments) when compared to the unrelated condition. Mixed design and within-subjects Analysis of Variance (ANOVA) tests were conducted to evaluate priming effects for speed. Logistic regression were used to evaluate the accuracy for L1 and L2 participants.

Results

Response Time Data

Descriptive analysis. A series of descriptive statistical analyses were conducted on the response time data to assess for outliers or errors. Descriptive statistics for the word raw response times for the L1 participants are presented in Table 2. Descriptive analyses demonstrated that the data for the hyponym-hypernym condition response times were normally distributed. However, the response time data for the hypernym-hyponym condition and the unrelated condition were not normally distributed. There was a single outlier in the hypernym-hyponym condition and two outliers in the unrelated condition. To correct the contaminated distribution and allow accurate interpretations through parametric statistics (Turkey, 1960), the data was first transformed using logarithmic formulas (logarithms to the base of 10; Field, 2005; Larson-Hall & Herrington, 2010). The logarithmic transformation corrected for one outlier in the unrelated condition. Not wanting to risk the independence of the data by removing the remaining outliers (Larson-Hall & Herrington), the scores for these outliers were transformed by assigning them a score that equaled the mean plus two standard deviations (Field).

Descriptive statistics for the word raw response times for the L2 participants are presented in Table 3. Descriptive analyses demonstrated that the response times for the unrelated condition were normally distributed. However, the response time data for the hyponym-hypernym condition, and the hypernym-hyponym condition were not normally distributed. There was a single outlier in the hyponym-hypernym condition and two outliers in the hypernym-hyponym condition. As with the L1 data, the data was first corrected for contaminated distribution through logarithmic formulas (logarithms to the base of 10; Field, 2005; Larson-Hall & Herrington-Hall, 2010). The logarithmic transformation corrected for all outliers in all the conditions.
Response Time Data

A mixed design ANOVA was conducted to assess overall effects of condition (hypernym-hyponym, hyponym-hypernym, and unrelated conditions) regardless of nativeness (i.e., L1 or L2 participants), differences in overall response times with respect to nativeness, and interactions between condition and nativeness. This ANOVA was followed by within subjects ANOVAs for L1 and L2 participants to examine differences in responses time between the conditions based on nativeness.

Mixed design ANOVA. There was a significant main effect for condition, $F(2, 164) = 12.477, p < .001, \eta^2 = .132$. Tukey contrasts revealed that responses times for all participants were significantly faster for the hyponym-hypernym condition than for the unrelated condition, $F(1, 82) = 28.335, p < .001, \eta^2 = .257$, and faster for the hypernym-hyponym condition than for the unrelated condition, $F(1, 82) = 6.918, p < .010, \eta^2 = .078$. There was a significant effect for nativeness indicating that response times for L1 participants were faster than response times for L2 participants, $F(1, 82) = 24.632, p < .001, \eta^2 = .231$. However, Levene’s tests for equality of error for all conditions between L1 and L2 participants were significant, indicating response times between L1 and L2 participants were not equal in variance. There was not a significant interaction effect between the conditions and the nativeness of the participant, $F(2, 164) = 1.254, p > .050, \eta^2 = .015$, indicating the response times between L1 and L2 participants were not significantly faster among the conditions.

Within-subjects ANOVA L1 participants. The within-subjects ANOVA of the response time data for the L1 participants showed a significant overall effect for
condition, \( F(2, 84) = 12.161, p < .001 \) \( h^2 = .225 \). Tukey contrasts revealed that responses times were significantly faster for the hyponym-hypernym condition than for the unrelated condition, \( F(1, 42) = 20.980, p < .001 \) \( h^2 = .333 \), and faster for the hypernym-hyponym condition than for the unrelated condition, \( F(1, 42) = 11.651, p < .001 \) \( h^2 = .217 \).

**Within-subjects ANOVA L2 participants.** The within-subjects ANOVA of the response time data for the L2 participants showed a significant overall effect for condition, \( F(2, 80) = 4.418, p < .050 \) \( h^2 = .099 \). Tukey contrasts revealed that responses times were significantly faster for the hyponym-hypernym condition than for the unrelated condition, \( F(1, 40) = 10.244, p < .010 \) \( h^2 = .204 \). However, response times were not significantly faster for the hypernym-hyponym condition when compared to the unrelated condition, \( F(1, 40) = 0.618, p > .050 \) \( h^2 = .015 \).

**Accuracy Data**

Because a ceiling effect existed for the accuracy data (i.e., participants could not receive a score above 100 percent accuracy and thus there was a predefined level above which variance could not be measured), no distribution corrections were made. Thus, all data points were analyzed as raw scores (see Table 5). The ceiling effect also means that the data is not normally distributed and ANOVAs could not be conducted. Instead, logistic regressions were conducted to assess differences between conditions and between L1 and L2 participants.

**All participants.** Accuracy scores significantly predicted the hyponym-hypernym condition from the hypernym-hyponym condition, \( \chi^2 = 7.556, df = 1, N = 84, p < .010 \) and the hyponym-hypernym from the unrelated condition, \( \chi^2 = 7.885 \),

<table>
<thead>
<tr>
<th>Table 5. Descriptive Statistics for Accuracy Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>Hyponym-hypernym condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Hypernym-hyponym condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Unrelated condition</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Accuracy scores did not significantly predict the hypernym-hyponym condition from the unrelated condition. This finding indicates that participants were more accurate in the hyponym-hypernym condition than in the other conditions.

**L1 participants.** Accuracy scores did not significantly predict any of the conditions indicating that participants were as accurate in all conditions.

**L2 participants.** Accuracy scores significantly predicted the hyponym-hypernym condition from the hypernym-hyponym condition, $\chi^2 = 12.494$, $df = 1$, $N = 41$, $p < .001$ and the hyponym-hypernym from the unrelated condition, $\chi^2 = 8.631$, $df = 1$, $N = 41$, $p < .010$. Accuracy scores did not significantly predict the hypernym-hyponym condition from the unrelated condition. This finding indicates that L2 participants were more accurate in the hyponym-hypernym condition than in the other conditions.

**Discussion**

This study has demonstrated that priming effects exist for both the hypernym-hyponym and the hyponym-hypernym conditions for L1 participants indicating that fluent lexical networks contain hierarchical clusters of related items that promote spreading activation between superordinate and subordinate terms and between subordinate and superordinate terms. However, for the L2 participants, the results indicate unidirectional priming effects suggesting that L2 lexicons are characterized by network connections that activate between subordinate and superordinate terms, but not between superordinate and subordinate terms. Such findings can provide important information about the organizational properties of L1 and L2 lexicons and the differences between fluent and developing lexicons.

For L1 speakers, the results provide additional evidence for the spreading activation model and the strength of hierarchical relations in the organization of the mental lexicon. In fluent lexicons, hyponyms activate hypernyms (i.e., the prime word car stimulates the target word vehicle) and hypernyms activate hyponyms (i.e., the prime word car stimulates the target word ambulance). This finding provides evidence that fluent lexicons can be characterized by organizational principles that allow for set inclusion among category members. These organizational features likely allow users to generalize terms because the terms are part of a conceptual category and are not isolated from one another. Such organizational principles provide for economic lexical representations and the subsequent priming effects reported by the response times in this study.
For advanced L2 speakers, the results suggest that lexical networks do not develop to the same degree as found in L1 participants and call into question the strength of L2 lexical networks and the use of spreading activation models to explain lexical associations between hierarchically related words in L2 mental lexicons. The results indicate that priming effects for L2 speakers in this study are not only unidirectional (only between a hyponym as a prime and a hypernym as a target), but, more importantly, the targets that shared fewer conceptual cues with the prime word, and were thus less similar, demonstrated activation. A spreading activation model would predict that concepts that were more similar would be stored more closely together and thus have more rapid activation. However, the findings of this study indicate that only hyponyms prime hypernyms, which contain fewer shared items with the prime. Conversely, hypernyms do not demonstrate priming effects for hyponyms even though hypernyms should share a greater number of cues with the hypernym prime and would thus be more similar (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). An example of this relationship is contained in the prime word *car*, which acts as a hyponym of *vehicle* and a hypernym of *ambulance* (see Figure 1). *Ambulance*, as a hyponym of *car* contains more shared items with *car* than *vehicle* does (e.g., four wheels, doors, a steering wheel, a windshield, seats, an internal combustion engine). *Vehicle* is less similar than a car and may or may not contain reliable cues (e.g., not all vehicles have wheels, doors, steering wheels, windshields, seats, and internal combustion engines). However, even though *vehicle* is less similar to *ambulance*, the results from the L2 participants support the notion that the word *car* will prime *vehicle*, whereas it will not prime *ambulance*.

That a spreading activation model does not explain L2 priming effects for hypernymic relations may be a result of dimensionality. For each hyponym, there is only one hypernym; however, for each hypernym there may hundreds of hyponyms (i.e., greater dimensionality). As a result, there may be a dispersion effect for hyponyms in that activation is spread across a greater number of items, thus increasing processing time and reducing the priming effect. No such dispersion effect would exist for hypernyms and, as a result, a stronger priming effect would occur. The results reported in this study may therefore be the result of the number of related conceptual links and not the distance between these concepts.

Figure 1. Immediate hierarchical relationships for the word car.
Alternatively, the reported findings may be the result of markedness in that a hyponym is semantically more complex than its hypernym because it contains more semantic features. As a result the hyponym also has less potential as a referent for its hypernym (Andersen, 2001; Jacobson, 1983; Kearns, 2006). Since marked forms are argued to be more difficult to learn, it is possible that L2 speakers have lower fluency and accuracy with hyponyms as compared to hypernyms, affecting the speed of word recognition. Such a position is supported by the accuracy data reported in this study, which indicated L2 participants were more accurate in the hyponym-hypernym condition than in the other conditions. Such a position is also supported by previous studies that showed that L2 learners produce words that are less specific as a function of time spent learning English (Crossley et al., 2009) and that L2 learners use more general words than specific words (Levenston & Blum, 1977).

Overall, this study provides some evidence to support the notion that hierarchical lexical networks are not fully developed in advanced L2 speakers to the same degree as found in L1 speakers. This is in contrast to past L2 studies investigating other aspects of lexical knowledge (i.e., collocations, antonyms, and synonyms; Dong et al., 2005; Frenck-Mestre & Prince, 1997), which have demonstrated priming effects that are similar to those found for L1 speakers.

Conclusion

This study has provided additional evidence for the existence of bi-directional hierarchical lexical networks in L1 lexicons. However, for advanced L2 speakers, such evidence was lacking suggesting that advanced L2 speakers may not access and store hierarchical lexical information in a similar manner to L1 speakers.

In general, the experimental designs used in this study provide confidence in the reported findings. However, the L2 RTs and the standard deviations for these RTs were greater than expected. Such variability in the L2 RTs, as compared to the L1 RTs means that statistical power between the two analyses was not equal and that a greater L2 sample size may be needed to provide appropriate power in order to generalize the findings across the population. Thus, the L2 results should be viewed with caution. Also, only subject analyses and not item analyses were not conducted for the data. This should also be considered a limitation. Demographically, this study used a generalized definition of L2 learner and did not control for age of learning a second language or for the amount of time spent in an environment wherein English was the dominant language. Future studies may want to control for age of exposure and learning environment, both of which may play a role in the development of lexical networks. Additionally, this study used a
within language design in which the prime and the target words were in English. It is possible that priming effects may exist between languages (i.e., a word in the L1 may prime a word in the L2) and not within languages.

Acknowledgments

The author would like to thank Christie Collins, Laura Varner, and Ashely Titak for their help in collecting participant data. The author is also indebted to Danielle McNamara, You Jin Kim and Kim McDonough for their assistance with the experimental design. Lastly, the author thanks the anonymous reviewers and the editors at the Mental Lexicon for their assistance in developing this paper.

Note

1. Alternative theories support the notion that categories contain only information that is common across all members (e.g., Collins & Quillian, 1969). In the examples provided in Figure 1, craft, car, and rocket would thus share has frame and missile, test rocket, and multi-stage rocket would share has jet engine, but not has frame.

References


Corresponding Address

Scott A. Crossley
Department of Applied Linguistic/ESL
Georgia State University
34 Peachtree St. Suite 1200
One Park Tower Building
Atlanta, GA 30303

Tel.: (404) 413-5179
E-mail: sacrossley@gmail.com