Frequency, neighborhood density, age-of-acquisition, lexicon size, neighborhood density and speed of processing: Towards a domain-general, single-mechanism account

Conference Paper · January 2007

1 author:

Vsevolod M Kapatsinski
University of Oregon
86 PUBLICATIONS 552 CITATIONS
SEE PROFILE

FREQUENCY, AGE-OF-ACQUISITION, LEXICON SIZE, NEIGHBORHOOD DENSITY, AND SPEED OF PROCESSING: TOWARDS A DOMAIN-GENERAL, SINGLE MECHANISM ACCOUNT*

VSEVOLOD KAPATSINSKI
University of New Mexico[1]

This paper proposes a domain-general, single-mechanism account of type and token frequency, neighborhood density, age-of-acquisition, speed of processing, target degradation, habituation, preexposure, and desensitization effects. Evidence from priming, word recognition, and associative learning is combined to provide support for Local Activation Spread Theory (LAST), which proposes that memory is a localist associative network that contains both type and token nodes where all types are interconnected while a token is only linked to one type.

1. PREVIOUS ACCOUNTS.

Within Network Theory, Moder (1992) proposed that high token frequency weakens a word's connections to neighboring words. Priming occurs by the spread of activation from the prime to the target. Since high-frequency words have weak connections, they receive less activation from their neighbors and their neighbors receive less activation from them. This account does not explain why high token frequency also reduces identity priming where the prime and the target are the same word.

Ratcliff and McKoon (1988) proposed that the prime and the target form a compound cue used to access long-term memory. The greater the familiarity of the cue, assessed as a weighted sum of the familiarities of the prime and the target, the faster long-term-memory access can occur. The greater the frequency of the target, the smaller the prime's contribution to overall familiarity of the cue, hence the amount of priming observed is smaller when the target is very familiar, i.e. has high token frequency. This account predicts that high prime frequency should increase priming, since high-frequency primes would contribute much to the overall familiarity of the cue, while in reality high prime frequency reduces the magnitude of priming observed.

Plaut and Booth (2000) have proposed a distributed connectionist model as a way to account for frequency effects. In this model, the prime and the target are overlapping patterns of activation (which can be understood as ordered sets of 1's and 0's) superimposed on the same set of nodes. The more similar the prime and the target are, the more values they will have in common, hence the transition from the prime to the target will be less costly for the network when the prime and the target are similar: fewer changes to node activation values would have to be made. Nodes have sigmoid activation functions, hence more input activation is required to change a node's activation level by a fixed amount in the direction of a value (1 or 0) when the node's resting activation level is already close to that value. Hence, activation of a high-frequency node value during prime presentation will not improve the node's ability to take on

---

* Many thanks to Joan Bybee, Rena Torres-Cacoullos, Catherine Travis, and the anonymous reviewers for helpful comments on earlier drafts of this paper and to Larry Gorbet, Dan Jurafsky, James MacFarlane, James McClelland, Jill Morford, Dan Olsher, David Pisoni, and the audience at HDLS VI for stimulating discussions of the issues involved.

[1] Currently at Indiana University

that value as much as when the value is mid-frequency. This account does not explain why prime frequency and target frequency influence the amount of priming. According to the account, it is only the frequency of the node values that the prime and the target share that should matter because the prime and the target do not have an independent existence. The importance of whole-unit frequency in priming is shown by priming asymmetries. For a given pair of stimuli, less priming is observed when the high frequency member of the pair is the prime than when it is the target in semantic (Koriat 1981, Chwilla, Hagoort, and Brown 1998), visual (Rueckl 2003), morphological (Schriefers, Frederici, and Graetz 1992, Feldman 2003), acoustic/phonetic (Goldinger, Luce, and Pisoni 1989) and phonological (Radeau, Morais, and Segui 1995) priming.

2. BASIC FEATURES.

2.1. ARCHITECTURE.

LAST proposes that memory is a localist associative network where every unit type - a word, a morph, a phone, a construction, a non-verbal stimulus – corresponds to a TYPE NODE, and every presentation of a type forms a TOKEN NODE, cf. Hintzman (1986). Evidence for the type/token distinction comes from several sources.

Moscoso del Prado Martin, Ernestus, and Baayen (2004) modeled the English past tense in a connectionist framework either presenting each present-past pairing the same number of times or the number of times proportional to the pairing's token frequency in the CELEX database. The network exhibited better performance on items it was not exposed to when the type-frequency based training regime was used. However, using a type-frequency-based training regime also leads to predicting same amounts of overgeneralization for regularized high- and low-frequency irregulars, while low frequency irregulars are actually regularized more often (Bybee 1995, 2001). Furthermore, using type frequency as a training regime leads to massive overgeneralization, since irregular forms are no longer frequent enough to withstand analogical leveling.

Albright and Hayes (2003) presented subjects with words that were highly similar to very frequent irregulars in hopes of obtaining analogies based on a single form. Their subjects almost never used the highly similar irregulars as analogical models, which was predicted by Albright and Hayes' Rule-Based Learner because of its bias in favor of rules that apply to more types. By contrast, when the present author trained a connectionist model over the same training corpus, using a token-based training regime, it found the high-frequency irregulars extremely attractive as analogical models. While for the subjects this was the experimental condition that most disfavored using an irregular, it was the most favorable context for irregular past tense formation for the connectionist model. Bybee (1995, 2001) and Moder (1992) have presented further evidence that type and token frequency have the opposite effects on morphological productivity.

Miller (1994) and Kapatsinski (2004) show that speakers store exemplar-specific information about words and phonemes they hear, such as their temporal duration, characteristics of the speaker's pronunciation, and so on, requiring token nodes. Miller found that speakers can reliably judge the goodness of different exemplars of a phoneme. Kapatsinski found that the same two words differing by a single segment are perceived to sound less similar to each other when the segment they differ by is artificially long than when it is artificially short, although the segment is phonologically identical in the two conditions, suggesting that phonological similarity judgments are based on comparing tokens. Finally, Hall (2003) has demonstrated that token-based activation leads to habituation while type-based activation retards it.

Evidence for types comes from McLennan, Luce, and Charles-Luce (2003), who found that allophonic variation has no effect on the amount of repetition priming observed. In particular, for English speakers, using nonsense primes and targets, it does not matter whether both the prime and the target contain an intervocalic [t] or either the prime or the target contains a flap in the same position. This finding suggests that the stimuli containing [t] and those containing the flap were not mapped onto different types, unlike primes and targets produced by different speakers (Goldinger 1992) or in different fonts (Tenpenny 1995, Goldinger, Azuma, Kleider, and Holmes 2003), which produce less priming than primes and targets produced in the same voice or font. Finally, the existence of long-term cross-modal morphological priming (Marslen-Wilson, Tyler, Waksler, and Older 1994) and syntactic priming in the absence of lexical overlap (Bock 1989, Bock and Loebell 1990) suggests the existence of rather abstract type nodes. Further evidence of abstract types is provided by the fact that reaction times are modeled well by log word frequency, rather than the sum of log frequencies of voice-specific word representations.

In LAST, all types are connected to each other, although these connections vary widely in their strength, while a token is linked to only one type. That is, a separate token node is created for every unit segmented out of the speech stream. Thus, presentation of a sentence may result in the formation of token nodes for all the constructions, prefabs, words, morphemes, syllables and phonemes.

Evidence for full connectivity between types comes from Ratcliff and McKoon (1981), who found that degree of semantic relatedness between the prime and the target influences the magnitude of the priming effect but not how soon after prime presentation the effect can be observed; thus, closely related words and less strongly related ones appear to be equally close to each other. If this were not the case, activation spreading from the prime would take longer to reach distantly related targets than closely related ones. Therefore, more time would need to pass since prime presentation for the effect to be observed with distantly related targets than with closely related targets. Since no differences are found, activation must reach distantly related and closely related targets simultaneously.

Earlier spreading activation theories (Anderson 1983) dealt with this finding by proposing that activation spreads so fast that its spread cannot be detected. This assumption is not made in LAST. This leads LAST to predict that identity priming should be observed at a shorter prime-target stimulus-onset-asynchrony than associative priming. This prediction remains to be tested. The architecture is presented in Figure 1.

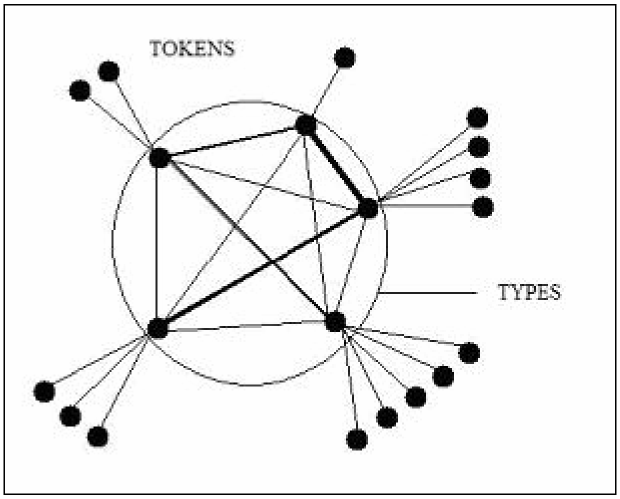

FIGURE 1. ARCHITECTURE OF MEMORY ACCORDING TO LAST[2]

2.2. LINK STRUCTURE.

In LAST, a LINK is a unidirectional channel of activation flow in that it only transmits activation from the HEAD of the link to its TAIL. Each CONNECTION in the network consists of two links such that the head of one link is the tail of the other and vice versa. Each link has a PROPAGATION FILTER (PF). The resting activation value (r-value) of a link's PF is directly proportional to how much activation is allocated to the link by its head. The PF, however, is not affected by activation spreading through the link it is located on (Sumida and Dyer 1992, Sumida 1997). If activation flowing through a link increased the PF's r-value, the link would strengthen whenever its head is activated, wrongly predicting that high token frequency words are better linked to their neighbors and therefore are better able to activate or prime them, against findings that high token frequency actually corresponds to reduced priming (see next section).

On the other hand, a link should strengthen whenever its head and tail are activated simultaneously (co-activated) to allow associative learning to occur. This would imply that the PF's r-value is raised whenever the head and the tail of the link are co-activated and hence that the PF of a link is a tail on a subsidiary link (LINKTRON) headed by either the head or the tail of the link the PF mediates. The influence of this linktron must, however, be counteracted whenever the head or the tail is activated in isolation. The following 12 structures are possible a priori (Figure 2). However, we will see in Section 2.4 that there are strong empirical constraints on possible structures.

2.3. DYNAMICS OF ACTIVATION SPREAD.

LAST proposes that each node has an ENTRANCE THRESHOLD and an EXIT THRESHOLD. After the entrance threshold is reached, activation can flow into the node. After the exit threshold is reached, activation is divided between the node itself and all links headed by the node. The amount of activation leaving a node is limited and as activation is leaving a node it is divided between all links connected to it. LAST differs from previous semantic networks models (e.g.,

---

[2] Lines represent connections, width of line represents connection strength, filled circles represent nodes. The circle surrounding the type nodes has no theoretical significance.

Anderson 1983, 2000, Neely 1991) in assuming that the node itself also participates in this competition so that the more links are connected to a node, the less activation will be allocated to any one link and to the node itself. Finally, LAST proposes the EQUITY PRINCIPLE, which states that the amount of activation allocated to a link is directly proportional to the strength of that link. However, while the strength of a link is equivalent to the resting activation level of its propagation filter, the strength of a node, lacking a propagation filter, is fixed. Activation stored in a node or a PF is assumed to decay as time progresses. The decay function is exponential or power-law based. Decay rate at a point of time is specific to the date of birth and size of a given ACTIVATION UNIT where an activation unit is a moving element defined by its current location as well as time and place of origin. Activation units created recently decay at a faster rate than those created long ago and the larger the activation unit, the slower its rate of decay. This size-dependent decay (SiDeD) hypothesis is unique to LAST and is the opposite of what is assumed by ART (Grossberg 1995).

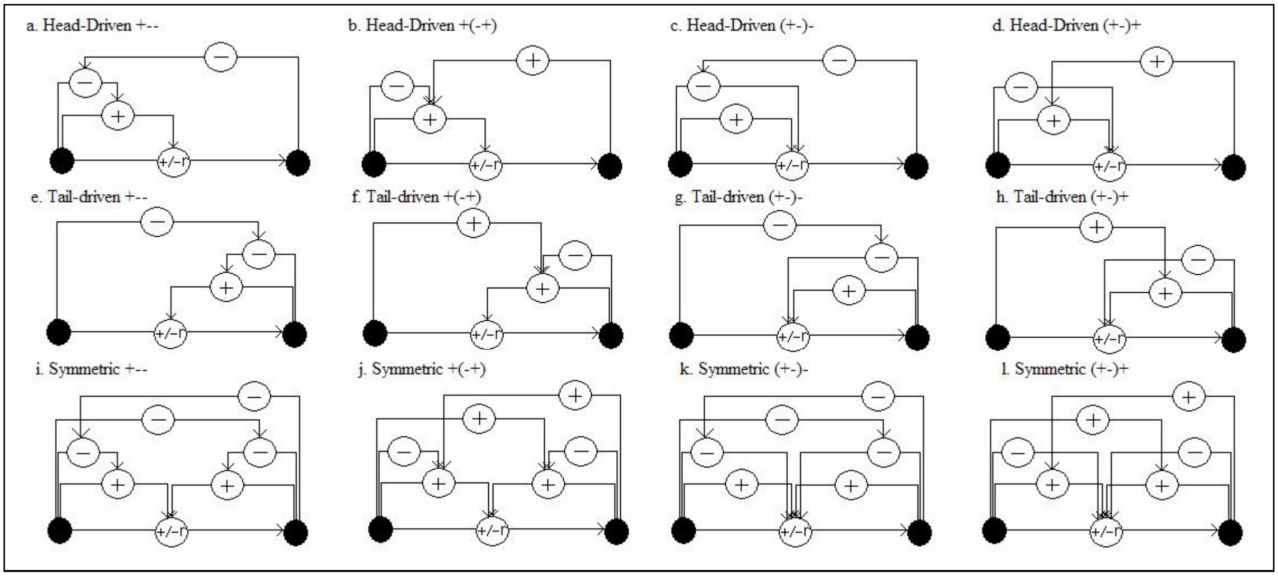

FIGURE 2. A PRIORI POSSIBLE LAST LINK STRUCTURES[3]

2.4. EMPIRICAL CONSTRAINTS ON LINK STRUCTURE.

If the structure of the link is symmetric, the strength of both links connecting two nodes is identical. If the structure is asymmetric, the DRIVING NODE of a link is the only one sending activation to the PF. A very robust finding is that if a stimulus A usually precedes stimulus B in a subject's experience, then priming would be much larger when A is the prime and B is the target than if the reverse is true (Koriat 1981, Chwilla et al. 1998). If A precedes B, B is at a higher level of activation when A and B become co-activated. This means that more activation will flow to the PF of the link driven by B. Since it is the A→B link that is strengthened more, it is A→B that is driven by B, hence links are tail-driven.

An increase in the token frequency of the driving node would mean less activation being

---

[3] Signs shown next to structure names indicate signs for PF r-values at progressively higher levels of embedding. Parentheses indicate same tail. If an r-value is negative, excitatory activation entering the node will become inhibitory when it leaves the node, where INHIBITORY activation reduces the r-value of nodes it enters.

sent to the PF's of links driven by the node since more activation would dissipate into the driving node tokens. Therefore, in a given connection, the link driven by the high-frequency node will, in absolute terms, be weaker than the link driven by the low-frequency node. For a given pair of stimuli, less priming is observed when the high frequency member of the pair is the prime than when it is the target in semantic (Koriat 1981, Chwilla et al. 1998), visual (Rueckl 2003), morphological (Schriefers et al. 1992, Feldman 2003), acoustic/phonetic[4] (Goldinger et al. 1989), and phonological (Radeau et al. 1995) priming. Since in priming activation must flow from the prime to the target, we may conclude that links are head-driven.

However, there is an alternative explanation for asymmetric frequency effects in priming: the size-dependent decay function that is also necessary to account for the longer duration of morphological and identity priming compared to phonological, semantic, and orthographic priming. If larger activation units decay more slowly, less priming should be observed when the prime is high frequency and the target is low frequency than when the prime is low frequency and the target is high frequency. The reason is that the major reduction in activation unit size occurs earlier in a unit's lifetime when the prime is high frequency than when the target is high frequency. As a result, activation spreading from the prime decays more before target presentation when the prime is high frequency than when the target is high frequency, resulting in decreased priming.

This account requires links headed by the high-frequency node to be absolutely stronger than those headed by the low-frequency node (because their PF's are driven mostly by the low-frequency node) but relatively weaker (because they have to compete with many type→token links). That is, the difference in node frequency must increase the amount of activation allocated to the driving linktron less than is decreases the amount of activation allocated to the link caused by the difference in node frequency, or $r_{Tt} > r_{linktron}$, that is, the amount of activation allocated to a type→token link must be greater than the amount of activation allocated to a linktron

This means that the learning of associations through link strengthening is a slow process, making one-shot learning the responsibility of unit/node formation. McClelland, McNaughton, and O'Reilly (1995) have reached the same conclusion based on the necessity for slow learning rates in connectionist networks.

Link structures provide predictions for associative learning.[5] The strength of an association after classical conditioning is measured by how likely the conditioned stimulus (CS) is to evoke the response typically associated with the unconditioned stimulus (US). Thus, the task responds to the strength of the connection leading from the CS to either the US or the response. If links were purely tail-driven, we would not expect the frequency of the CS to influence the strength of the CS→US or CS→response link, contrary to the findings (CS preexposure effect, Hall 2003). Thus, we must assume that, while the linktron headed by the tail of the link is stronger than the linktron headed by the head of the link[6], both exist, thus only i-l are possible structures.

Linktron PF's, unlike link PF's, are non-trainable. If this were not the case, the linktron headed by one of the linked nodes would strengthen or weaken one of the linktrons headed by

---

[4] The fact that inhibitory priming is reduced by high frequency as well as excitatory priming means that the effect cannot be accounted for by a ceiling effect.

[5] This is not to say that according to LAST language acquisition is simply associative learning. A hallmark of language acquisition is one-shot learning, which in LAST is accomplished by the formation of a new node.

[6] Note that this does not require the frequency of the US to have a greater influence on link strengthening than the frequency of the CS because the CS may be connected to the response directly and the US may have a strong connection to the response because of its intrinsic content.

the other node, upsetting the equilibrium and allowing the other node to strengthen the link when activated in isolation. This would wrongly predict that high token frequency increases associability and priming as long as the stimulus has ever been co-activated. If linktron PF's are non-trainable, (+-)+ structures are impossible because the linktron headed by the non-driving node has no effect. Therefore, only i, j, and k are possible structures.

Linktrons are fixed to have the same strength as the linktrons whose PF's they influence. That is, a linktron can be either able or unable to transmit activation.

Finally, we should note that for activation to spread from a node on a link to the PF of the link before activation from the head reaches the link's PF through the link, linktrons must be able to transmit activation faster than links, and linktrons whose tails are linktron PF's must be faster still.

3. THE LAST ACCOUNT OF THE FINDINGS.

In priming, when the prime is presented, matching types are partially activated. Whenever a type passes the exit threshold, a token node is created storing the information about that particular instance of the type, and activation starts to spread from the type node. Since the spread of activation from the node along type-token links occurs only after it has been activated, speed of the type node's recognition is directly proportional to its r-value: the lower the type node's r-value, the more input activation is needed to reach the level sufficient for recognition. Thus, the greater the token frequency of a type, the faster its recognition should occur. This is precisely what has been found by, e.g., Coltheart, Davelaar, Jonasson, and Besner (1977), Glanzer and Ehrenreich (1979), Gordon (1983), Norris (1984), Goldinger et al. (1989), Luce, Goldinger, Auer, and Vitevitch (2000), and Plaut and Booth (2000). Slow recognition feeds back to low experienced token frequency since slow processors would not recognize as many tokens of a type present in the environment as fast processors.

Once the prime type is activated, activation starts to flow out from it. As it is leaving the node, it is divided between the node itself and all links headed by the node, one of the links being tailed by the target, which has not yet been presented. Thus, the greater the number of links headed by the prime's type, the less activation will remain in the node and the less activation will be allocated to any one link. Given that every type is connected to all other types while a token is only connected to one type, the only factors influencing the number of links radiating from a type are its token frequency and the number of types in the lexicon. Therefore, the higher the token frequency of the prime, the less priming, including identity priming, should occur as found by, e.g., Scarborough, Cortese, and Scarborough (1977), Forster and Davis (1984), Norris (1984), Stark (1997), Perea and Rosa (2000), and Versace and Nevers (2003) for identity priming, by Moder (1992) for morphological priming, by Thomsen, Lavine, and Kounios. (1996) for semantic priming, and by Goldinger et al. (1989) and Luce et al. (2000) for phonological priming.

In addition, the smaller the size of the lexicon, the more activation is allocated to any one node, hence more priming in children and late signers (as found by, e.g., Perfetti and Hogaboam 1975, Simpson and Lorsbach 1983, Emmorey, Corina, and Bellugi 1995, Nation and Snowling 1998, Castles, Davis, and Letcher 1999), and faster recognition (and higher rated familiarity) of words learned early in life compared to words with the same token frequency learned later in life, indicating higher resting activation levels for early-acquired words (as found by Newman and German 2002, Zevin and Seidenberg 2004, and Ghyselinck, Lewis and Brysbaert 2004).

The finding that earlier-acquired words exhibit less priming (Barry, Hirsh, Johnston, and WIlliams 2001) is predicted because when the lexicon is small more activation will reach the propagation filter of a link when the nodes it links are co-activated. Thus, LAST predicts that early age-of-acquisition should be correlated with having many strong associates, as found by Steyvers and Tenenbaum (2005) for semantic associates.

Since slow processing leads to low token frequency, more priming should also be observed in slow word-recognizers when lexicon size is controlled, as found by Plaut and Booth (2000). The idea of type-token links is further supported by the finding that direct perceptual exposure to a type, which increases both its token frequency and resting activation level, leads to habituation, or less activation spreading to neighboring types, while associative activation received from another type leads to dishabituation because less activation is necessary to reach the exit threshold (Hall 2003).

Low frequency orthographic primes produce no inhibitory priming while high frequency orthographic primes significantly inhibit their targets (Drews and Zwitserlood 1995, Brysbaert, Lange, and van Wijnendaele. 2000). On the other hand, Goldinger et al. (1989) and Luce et al. (2000) observed that high-frequency primes inhibit targets less than low-frequency primes do in phonological priming. This pattern of results is accounted for by proposing that the function of amount of priming[7] by prime or target frequency is the same, regardless of modality (Figure 3). However, orthographic representations are acquired later than phonological representations, at a time when the lexicon is larger. Therefore, connections between orthographic representations will be weaker than those between phonological ones, since less activation will reach the PF's of links connecting orthographic representations after the same number of co-activations. Consequently, less activation/inhibition from the prime will reach the target in orthography resulting in greater incidence of associative activation not being let into the target node by the entrance threshold.

Lack of sensitivity to word frequency can also be observed if the priming is sublexical, occurring between parts of the prime and the target rather than between prime and target themselves. Radeau et al. (1995) review the literature on facilitatory phonological priming, which is based on word-final overlap, and finds that the priming is not sensitive to word frequency, indicating its sublexical locus. With sublexical priming, LAST predicts that the amount of priming should be influenced by the frequency of the sublexical units involved and not word frequency.

A finding inconsistent with the LAST account of priming (but predicted by the activation functional account of Plaut and Booth 2000) would be if low frequency primes produced some priming but the amount of this priming were less than that produced by high frequency primes. No such evidence exists at present (both prime frequency[8] and target frequency, cf. Schuberth and Eimas 1977, Neely 1991, Perea and Rosa 2000, exhibit an inverse correlation with amount of priming).

---

[7] By 'amount of priming' we mean the absolute magnitude of the change in the dependent variable (reaction time, brain wave latency and amplitude, error rate) due to related prime presentation compared to an unrelated prime.

[8] See citations above.

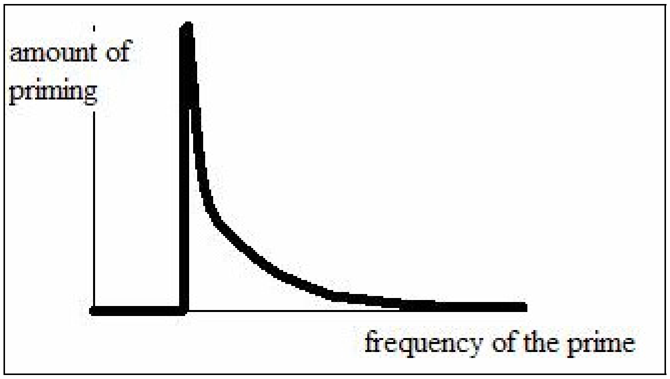

FIGURE 3. THE EFFECT OF THE FREQUENCY OF THE PRIME ON AMOUNT OF PRIMING HYPOTHESIZED BY LAST

Unlike similarity-based priming, identity priming can be observed with nodes with a very low r-value, since the input activation does not dissipate into the network before the prime type reaches the exit threshold and the entrance threshold is lower than the exit threshold. Stark (1997) has demonstrated that even visually presented pronounceable non-words produce identity priming and that, furthermore, they produce even more identity priming than low-frequency words.

LAST's Equity Principle states that the amount of activation allocated to a link is proportional to its strength (cf. Anderson 2000). If this is the case, we would predict that words that are semantically, phonologically, or orthographically similar to many other words, that is, words located in dense neighborhoods, should exhibit less priming than words located in sparse neighborhoods, since a link or node of a given strength will receive less activation in a dense neighborhood than in a sparse neighborhood. This is indeed what is found by Thomsen et al. (1996) for semantic priming and by Perea and Rosa (2000) for orthographic priming. In addition, since the source node's r-value is fixed, late-acquired words will exhibit more identity priming (since the source node would compete with weaker links) as found by Barry et al. (2001).

Even more direct evidence for the Equity Principle is provided by Anaki and Henik (2003), who find that if the target is given as an associate of the prime by a certain percentage of subjects in a free association task, there will be more priming between the prime and the target if the other associates of the prime are given by lower percentages of the subjects. For instance, suppose that 'mouse' is given as an associate of 'cat' by 50% of the subjects and 'nail' is given as an associate of 'hammer' by 50% of the subjects. Then if 'dog' is the next most popular associate of 'cat', produced by 45% of subjects, while 'tool' is the next most popular associate of 'hammer', produced by 10% of subjects, 'cat' will prime 'mouse' less than 'hammer' will prime 'nail' because the 'cat-mouse' connection has to compete for 'cat's activation with other strong connections while the 'hammer-nail' connection competes for 'hammer's activation only with weak connections.

The Equity Principle also explains the blocking effect in associative learning: if a US is already strongly associated with a CS (CS1), it is hard to associate with another CS (CS2) (Kamin 1969). This is because the linktron headed by the US and tailed by the PF of the US-CS2 link has to compete with a strong US-CS1 link, relative to the case in which the US has no strong associates. Consequently less activation will be allocated to the US-CS2's PF when US is strongly associated with CS1, leading to slower strengthening of US-CS2. Similarly, high-frequency CS1's and US1's are harder to associate with a CS2 or US2 because the PF's of links being acquired have to compete with many type-token links headed by CS1 or US1 (what are known as the CS preexposure effect and the US desensitization effect, cf. Hall 2003 for a review).

A related phenomenon is proceduralization, where parts of a frequently practiced sequence of actions become strongly associated as a result of practice while the verbal descriptions of the actions become inaccessible from the action (Ryle 1949).[9] Since the actions in the sequence develop strong connections, less activation would spread from nodes corresponding to any one of the actions to associates that are not practiced every time the action is performed, e.g., the verbal descriptions of the actions.[10]

Further evidence for the Equity Principle in associative learning comes from the interference paradigm of Barnes and Underwood (1959) and McGovern (1964) who found that when subjects are asked to learn a list of A-C stimulus pairs after learning a list of A-B pairs, they can recall B when presented with A worse than subjects who are asked to learn C-D pairs after learning A-B pairs.

The last issue for which the Equity Principle is relevant is neighborhood effects in word recognition. The effects appear to be inconsistent: words from dense neighborhoods are found to be recognized more slowly (e.g., Goldinger et al. 1989, Luce and Pisoni 1998, Vitevitch and Luce 1998, Luce et al. 2000) or more quickly (e.g., Andrews 1989, Forster and Shen 1996, Huntsman and Lima 2002). This inconsistency is in fact predicted by LAST, because high density leads to more associative activation reaching the type node corresponding to the word to be recognized but to a lower proportion of both associative and direct (token-based) activation sticking to the node to raise its r-value. Thus, the system is unstable and effects are inconsistent.

Nevertheless, low-frequency words should be more likely to show facilitatory effects of neighborhood density because their resting activation level depends less on direct activation, for which high density has only inhibitory effects. This proposal is supported by evidence obtained by Metsala (1997), who found that high density has facilitatory effect for low-frequency words, while it has an inhibitory effect for high-frequency words. Indirect support for the hypothesis is also provided by the fact that most studies providing support for inhibitory density effects use acoustic stimuli, and acoustic stimuli have a higher r-value than orthographic stimuli due to a higher token frequency and earlier age of acquisition.[11]

A challenge to any single-mechanism account of priming is to explain why morphological and identity priming persist over several intervening items, while semantic, phonological, and orthographic priming decay rapidly (cf. Stockall 2004), although long-term semantic priming has been observed by Becker, Moscovitch, Behrmann, and Joordens (1997). LAST handles this dissociation by observing that the activation units reaching morphologically related neighbors are larger, and much activation also stays in the node from which activation begins to spread. Morphological priming involves identity priming of roots, which are much less frequent than the phonological units that are identity-primed in phonological priming (cf. Radeau et al. 1995). Because of the relatively low frequency of the source node, little activation

---

[9] See Whittlesea (2003) on evidence contradicting ACT's (Anderson 1983) view that the procedural/declarative distinction is a distinction between cognitive modules.

[10] The Equity Principle also explains transitional probability effects without postulating that speakers calculate probabilities: a word that is very probable is a word whose connection to the previous word in the processing sequence is strong relative to its competitors.

[11] However, facilitatory density effects may in reality be sublexical frequency effects or the effects of the density and ease of access of competitors.

dissipates into its tokens, leaving much activation in the source node and allowing much activation to spread to the target. Support for this notion comes from observing no priming with high-frequency shared morphemes, i.e. inflectional affixes (Emmorey 1989). Syntactic priming may be thought of as identity priming of schematic constructions (Goldberg 2006). Since in LAST, the larger the activation unit, the slower the rate of decay, we predict that identity priming and priming that relies on stronger connections should decay less rapidly.

Lateral inhibition predicts that words whose neighbors are neighbors of each other should be easier to recognize than words whose neighbors are not neighbors of each other. Altieri (2006) found that the exact opposite is the case. These results are inconsistent with lateral inhibition and consistent with spreading activation.

Phonological priming of word beginnings is always inhibitory when strategic factors[12] are controlled (Goldinger et al. 1989, Radeau et al. 1995, Luce et al. 2000). On the other hand, phonological priming of word ends is excitatory (Radeau et al. 1995).

While this finding is standardly explained by lateral inhibitory connections (McClelland and Rumelhart 1981), LAST suggests an alternative mechanism. A new phonological token is created whenever a word is presented auditorily but the semantics is activated associatively and formation of a new token does not result. Whenever a part X is activated when a whole AX is presented, a new token of X is formed. Therefore, less activation will spread to wholes containing X from X in the future. This would not apply to the whole containing X present in the environment (AX) since X and AX would be co-activated strengthening the X-AX connection and thereby, via the Equity Principle, additionally reducing the amount of activation that would reach other wholes containing X from X in the future, leading to slower recognition times for those wholes. Hence, we observe inhibitory priming for phonologically but not semantically similar items. This does not apply to word-end-based priming because word ends are usually not necessary for word recognition. Rather, the priming that occurs reflects simply faster access to the sublexical units in the word's end. A recent experiment with sound similarity judgments has shown that words that share frequent phonemes are perceived to sound less similar than words that share rare phonemes, supporting the idea that connections are stronger between words that share parts that occur in few, as opposed to many, words (Kapatsinski 2006).

Bybee (1995) and Albright and Hayes (2003) propose that high type frequency makes morphemes more productive. In LAST, this is predicted to occur: since a novel word is not associated with any of the competing morphemes when presented, its presentation results in the activation of words that are similar to the stimulus. Activation spreads through the type nodes corresponding to those words, activating the associated morphemes. The larger the type frequency of a morpheme, the greater the number of possible mediators that would allow activation from the novel word to spread to the morpheme. On the other hand, the higher the token frequency of the mediators, the more activation will dissipate into their tokens and hence the less activation will reach the associated morpheme, hence lower productivity of morphemes with high token/type ratio (Bybee 1995).

In production, high-frequency words are observed to show greater gestural overlap and coarticulation (Hooper 1976, Pagliuca and Mowrey 1987, Browman and Goldstein 1990, Bybee 2001:69-85). If articulatory targets are the relevant types (Browman and Golstein 1990), and activation flows from the preceding to the following target (Pulvermuller 2003), the following target will get activated faster in a high-frequency word, or a high-probability target sequence, leading to gestural compression.

---

[12] I.e., factors that are under the goal-oriented control of the subject.

Finally, it is often observed that frequency effects on reaction time are larger in language comprehension than in language production (e.g., Jescheniak and Levelt 1994). We would like to propose that this is because the incoming stimulus is matched to tokens, rather than types, such that the amount of activation allocated to an existing token depends solely on the degree to which the token matches the incoming stimulus. For a high-frequency type, many tokens would match the incoming stimulus, resulting in more activation flowing into the type. In production, there is no need to access the tokens, hence perception involves one extra frequency-sensitive stage in which frequency effects are facilitatory.

The existence of token units connected to the type node allows the type to carry no content, being defined by the weighted average of the contents of the associated tokens, hence allowing type content to drift gradually (cf. Pierrehumbert 2001 for a model of sound change along these lines). However, Baayen, Levelt, Haverman, and Desserjer (2005) have recently demonstrated that word frequency effects in picture naming are excitatory in immediate naming but inhibitory in delayed naming. This effect is predicted if naming can be accomplished by simply accessing the type, rather than accessing all the tokens. Activating a type is easier when the type is high-frequency because high-frequency types have higher resting activation levels, producing the facilitatory frequency effect in immediate naming. However, as the time from initial type activation increases, activation starts to leak out of the word type node into the tokens, producing inhibitory frequency effects in delayed naming. These results provide converging evidence for the possibility of producing a word without token access.

4. FUTURE DIRECTIONS.

Link structures where two linktrons of opposite signs have the same tail but different heads (b and j in Figure 2) are unstable with the effect of the structure on the link's PF depending on, in part, the frequencies of the heads of the linktrons. Therefore, if a link strengthens whenever its head and tail are co-activated and not only when the co-activation occurs at a time when the head is less frequent and/or lower in neighborhood density than the tail, only i and k are possible asymmetric structures. Structure j displays a very interesting dynamic: whichever node is higher in frequency or density than the other loses the battle to control the linktron PF's tailed by linktrons with opposite-sign PF's. Since a node always inhibits its own excitatory linktron tailed by the link's PF and excites the other node's linktron leading to the link's PF, the higher-frequency/higher-density node is the only one that spreads activation to the PF of the link and thus is the only node responsible for strengthening the link. Therefore, structure j predicts that strength of the link is determined by whichever of the linked nodes is higher in frequency and/or higher in density. For i and k, both nodes influence the link's strength.

A problematic finding for the Equity Principle has been obtained by Zeelenberg (1998), who primed either A-C associations as well as A-B associations or only A-B associations. No effect of A-C presentation on the amount of priming produced by A-B presentation was observed, while the Equity Principle would predict that priming A-C should reduce A-B priming. However, the primed A-C associations were already strong, reducing the magnitude of the manipulation. More work is needed on this issue.

Another finding that is problematic for LAST is Plaut and Booth's (2000) observation of lack of a word frequency effect on semantic priming in poor readers who also exhibited more priming than controls matched for vocabulary size and exhibited a frequency effect in speed of word recognition. This finding is predicted by Plaut and Booth's (2000) model with the assumption that the poor readers fall onto the linear portion of the activation function. A possible explanation is that these subjects are more likely to forget exemplars with low resting activation levels than good readers. While high-frequency words are connected to more tokens in the normal system, each of the tokens has a lower-resting activation level than the tokens of low-frequency types do, and hence is more susceptible to forgetting. This requires the increased opportunities for getting activation caused by high type frequency to not provide the token with full compensation for decrease in size and consequent increase in decay rate of activation units received on each occasion.

Finally, a problematic result has been reported by Murray and Forster (2004) who found that frequency rank order provides a better correlation with reaction time and accuracy in the lexical decision task than does log frequency, providing an argument for a model of lexical access based on serial search. A difference between log frequency and frequency rank order is that log frequency is absolute while frequency rank is relative. Given the existence of competition for recognition, a fairer comparison would involve log frequency relative to competitors and frequency rank relative to competitors.

Unlike in distributed connectionist models, which conceive of similarity as overlap between patterns of activation (e.g., Plaut and Booth 2000), in LAST similarity could strengthen connections between nodes, just as co-occurrence does, through co-activation. When two units share a part, the activation of the part will strongly activate both of the units containing it. Crucially, activation from the part will reach the wholes that contain it simultaneously, resulting in co-activation of the wholes. If two wholes share many parts, they will be co-activated strongly and often and hence will develop strong connections. However, similarity effects are observed even with nonce stimuli that the subject has not perceived before (e.g., Albright and Hayes 2003). This means that similarity relations, unlike co-occurrence relations should be available as soon as a new node is created, which suggests that similarity should in fact be conceived as overlap.

McClelland et al. (1995) suggested that the brain combines localist and distributed representation, such that the hippocampus and the superficial layers of the neocortex contain localist representations, which are linked to distributed representations in the deeper layers of the neocortex. We would like to suggest that the LAST type and token nodes form the hippocampal/superficial neocortex system, and are linked to distributed patterns of activity in deep neocortex. That is, a type node or a token node specifies a sequence of activity in the distributed system (for ways to specify sequences in a distributed system, see Pulvermuller 2003).

The size-dependent decay (SiDeD) hypothesis predicts that priming should decay more slowly with low-frequency, low-density primes and targets than with high-frequency, high-density ones. Furthermore, slower priming decay should be observed in small lexicons. In general, the greater the magnitude of priming at the minimum stimulus-onset-asynchrony at which priming can be observed, the slower its decay should be. We are currently working on an experiment to test this prediction using identity priming.

Hay (2001) found that affixes that tend to derive words that are more frequent than their bases (high-relative-frequency words) also tend to be less productive. This finding is straightforwardly explained if morphemes and words compete for activation during word recognition, such that if a word is recognized faster than its constituent morphemes, the morphemes' resting activation level is increased less than when the morphemes are recognized faster than the word. This is expected in LAST because when the morphemes are recognized faster than the word, they receive direct activation due to token formation, while when the word is faster, the morphemes are activated associatively. Such a process would lead morphemes occurring in high-relative-frequency words to have a lower resting activation level than morphemes occurring in low-relative-frequency words. A prediction following from the account of the relative frequency effect offered above is that bases that occur in high-relative-frequency words should exhibit more identity priming, since they have fewer tokens associated with them. This prediction awaits testing.

Hay's results show that a full account of frequency effects must also include a model of segmentation. That is, how large and abstract are the units denoted by nodes. In LAST, a type is a unit that is sufficiently abstract for subjects to be sensitive to its frequency of occurrence and to be able to learn associations between it and other types. Frequency alone is insufficient to establish that some unit is stored as a node, since complex unit frequency can always be reduced to the frequencies of the parts and transitional probabilities between them. Minimally, learning an association between two existing nodes should be easier than creating a new node and associating it with an existing node. Based on this intuition, we have developed XOR learning, a general method of testing unithood that can also distinguish LAST from distributed models.

In XOR learning, subjects learn that stimuli A and B on their own are associated with a response, X, while A and B together are associated with a different response, Y. Thus, the subjects are forced to learn associations for the complex unit AB that could not be inherited from and are in contradiction with the associations of the unit's parts. If AB is already stored as a unit, it should be associated with Y more easily than if only A and B are stored.

If high frequency of co-occurrence can result in the formation of a complex unit, as predicted by LAST and other localist theories (Bybee 2002, Solan, Horn, Ruppin, and Edelman 2005), learning the AB-Y association should be easier when A and B co-occur than when they do not, controlling for the overall frequency of A and B. However, if high frequency of co-occurrence simply strengthens the connection between A and B, learning of the AB-Y association should be harder when A and B co-occur. If A and B are associated due to co-occurrence, presentation of A would activate B. In XOR learning, B is associated with a response incompatible with AB's response and would therefore interfere with the learning of AB-Y associations. Thus, XOR learning allows us to directly compare localist and distributed theories of mental representation. Even multi-layer connectionist networks, which can solve XOR problems, do not predict that co-occurrence of A and B should make predicting Y's occurrence easier.

5. CONCLUSION.

The basic idea behind this paper is to take a single mechanism: competition for a limited supply of spreading activation and see how much it can account for and what architectural assumptions must be made. The result, LAST, provides an explicit, testable account of frequency, neighborhood density, lexicon size, age-of-acquisition, target degradation, and speed of processing effects across tasks, domains, and species. We hope to have shown how bringing together evidence from different domains constrains a theory and makes previously unseen connections apparent and to have provided a range of testable hypotheses that would stimulate research in the area.

APPENDIX: INEVITABILITY OF LINK STRENGTHENING DUE TO COACTIVATION

This appendix answers the question of what needs to hold for links to be strengthened by coactivation of the linked nodes despite the addition of type→token links to each of the linked nodes. Strengthening in LAST means that the link AB will attract more activation ACT_{AB} after coactivation of A and B.

Before coactivation,

$$(1)\ ACT_{AB} = \frac{ACT_A \times r_{AB}}{\sum rn}$$

where r_{AB} is the r-value of the link AB, ACT_A is the activation that node A has prior to division, and

$$(2)\ \sum rn = r_T + r_{Tt} \times n_{Tt} + r_{\mu_{TT}} \times n_{TT} + r_{linktron} \times n_{linktron}$$

or the sum of strengths of the r-values of all links headed by A plus the fixed strength of the node.

After coactivation,

$$(3)\ ACT_{AB} = \frac{ACT_A \times (r_{AB} + \frac{ACT_B \times r_{linktron}}{\sum rn_b} + \frac{ACT_A \times r_{linktron}}{\sum rn_a})}{\sum rn + r_{Tt} + \frac{ACT_A \times r_{linktron}}{\sum rn_a}}$$

For the desired result, (3) must be larger than (1), or, assuming, for simplicity, ACT_A=ACT_B

$$(4)\ \frac{ACT_A \times r_{AB}}{\sum rn_a} - \frac{ACT_A \times (r_{AB} + \frac{ACT_B \times r_{linktron}}{\sum rn_a} + \frac{ACT_A \times r_{linktron}}{\sum rn_a})}{\sum rn_a + \frac{ACT_A \times r_{Tt}}{\sum rn_a} + \frac{ACT_A \times r_{linktron}}{\sum rn_a}} < 0$$

After reducing to a common denominator and simplifying that $\sum rn_b = \sum rn_a$ and ACT_A=ACT_B,

$$(5)\ r_{AB} \times \sum rn + r_{AB} \times \frac{ACT \times r_{Tt}}{\sum rn} + r_{AB} \times \frac{ACT \times r_{linktron}}{\sum rn} - r_{AB} \times \sum rn - \frac{ACT \times r_{linktron}}{\sum rn} \times \sum rn - \frac{ACT \times r_{linktron}}{\sum rn} < 0$$

$$(6)\ r_{AB} \times r_{Tt} \times \sum rn + r_{AB} \times r_{linktron} \times \sum rn - 2 r_{linktron} \times (\sum rn)^2 < 0$$

Thus, we receive

$$(7)\quad r_{AB} \times (r_{Tt} + r_{linktron}) < 2 r_{linktron} \times \sum rn$$

Since the sum of all r-values is a very large number, relative to any given r-value, this is guaranteed to be true in every network. Thus, the model can produce the desired behavior without any special parameter setting.

REFERENCES [13]

ALBRIGHT, A. and B. HAYES. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90.119-61.

ALTIERI, N. 2006. Clustering coefficient and variable processing of lexical items. Ms. Indiana University.

ANAKI, D. and A. HENIK. 2003. Is There a "Strength Effect" in Automatic Semantic Priming? *Memory and Cognition*, 31.262-72.

ANDERSON, J.R. 2000. *Cognitive Psychology and Its Implications*. New York: Worth.

ANDERSON, J. R. 1983. *The Architecture of Cognition*. Cambrudge, MA: HUP.

ANDREWS, S. 1989. Frequency and Neighborhood Effects on Lexical Access: Activation or Search? *JEP: LMC*, 15.802-14.

BAAYEN, R. H., W. J. M. LEVELT, A. HAVERMAN, and G. DESSERJER. Producing singular and plural nouns: evidence for morphological paradigms. Ms. Max Planck Institute for Psycholinguistics, Nijmegen.

BARNES, J. M., and B. J. UNDERWOOD. 1959. 'Fate' of First-List Associations in Transfer Theory. *JEP*, 58.97-105.

BARRY, C., K. W. HIRSH, R. A. JOHNSTON, and C. L. WILLIAMS. 2001. Age of Acquisition, Word Frequency, and the Locus of Repetition Priming of Picture Naming. *JML*, 44.350-75.

BECKER, S., M. MOSCOVITCH, M. BEHRMANN, and S. JOORDENS. 1997. Long-term semantic priming: A computational account and empirical evidence. *JEP: LMC* 23.1059-82.

BOCK, K. 1989. Closed-class immanence in sentence production. *Cognition* 31:163-86.

BOCK, K., and H. LOEBELL. 1990. Framing sentences. *Cognition* 35:1-39.

BOWERS, J. S., and C. J. MARSOLEK (eds.) 2003. *Rethinking Implicit Memory*. New York: OUP.

BROWMAN, C. P., and L. M. GOLDSTEIN. 1990. Tiers in Articulatory Phonology, with Some Implications for Casual Speech. *Papers in Laboratory Phonology* 1.341-76.

BRYSBAERT, M., M. LANGE, and I. VAN WIJNENDAELE. 2000. The Effects of Age-of-Acquisition and Frequency-of-Occurrence in Visual Word Recognition: Further Evidence from the Dutch Language. *European Journal of Cognitive Psychology*, 12.65-85.

BYBEE, J. L. 2002. Sequentiality as the Basis of Constituent Structure. In *The Evolution of Language out of Pre-Language,* ed. by T. Givón and B. Malle. Amsterdam: Benjamins.

BYBEE, J. L. 2001. *Phonology and Language Use*. Cambridge, UK: CUP.

BYBEE, J. L. 1995. Regular Morphology and the Lexicon. *LCP*, 10.425-55.

CASTLES, A., C. DAVIS, and T. LETCHER. 1999. Neighbourhood Effects on Masked Form Priming in Developing Readers. *LCP*, 14.201-224.

CHWILLA, D. J., P. HAGOORT, and C. M. BROWN. 1998. The Mechanism Underlying Backward Priming in a Lexical Decision Task: Spreading activation versus Semantic Matching. *QJEP*, 51A.531-60.

COLTHEART, M., E. DAVELAAR, J. T. JONASSON, and D. BESNER. 1977. Access to the Internal Lexicon. In *Attention and Performance VI*, ed. by S. Dornic. London: Academic Press.

---

[13] Here, JEP=Journal of Experimental Psychology (LMC = Learning, Memory, and Cognition, HPP = Human Perception and Performance), JPR=Journal of Psycholinguistic Research, JML= Journal of Memory and Language, JVLVB=Journal of Verbal Learning and Verbal Behavior, PP = Perception and Psychophysics, QJEP = Quarterly Journal of Experimental Psychology, LCP = Language and Cognitive Processes, PR=Psychological Review

DREWS, E., and P. ZWITSERLOOD. 1995. Effects of Morphological and Orthographic Similarity in Visual Word Recognition. *JEP: HPP*, 21.1098-1116.

EMMOREY, K. D. 1989. Auditory morphological priming in the lexicon. *LCP*, 4.73-92.

EMMOREY, K., D. CORINA, and U. BELLUGI. 1995. Differential Processing of Topographic and Referential Functions of Space. In *Language, Gesture, and Space*, ed. by K. Emmorey and J. Reilly, 43-62. Hillsdale: Erlbaum.

FELDMAN, L. B. 2003. What the repetition priming methodology reveals about morphological aspects of word recognition. In Bowers and Marsolek (2003), 124-38.

FORSTER, K. I. and C. DAVIS. 1984. Repetition Priming and Frequency Attenuation in Lexical Access. *JEP: LMC*, 10.680-98.

FORSTER, K. I., and D. SHEN. 1996. No Enemies in the Neighborhood: Absence of Inhibitory Neighborhood Effects in Lexical Decision and Semantic Categorization. *JEP: LMC*, 22.696-713.

GHYSELINCK, M., M. B. LEWIS, and M. BRYSBAERT. 2004. Age of Acquisition and the Cumulative-Frequency Hypothesis: A Review of the Literature and a New Multi-Task Investigation. *Acta Psychologica*, 115.43-67.

GLANZER, M., and S. L. EHRENREICH. 1979. Lexical Access and Lexical Decision: Mechanisms of Frequency Sensitivity. *JVLVB*, 18.381-98.

GOLDBERG, A. 2006. *Constructions at Work: The Nature of Generalizations in Language*. New York: OUP.

GOLDINGER, S. D. 1992. Words and Voices: Implicit and Explicit Memory for Spoken Words. Ph.D. Dissertation, Indiana University.

GOLDINGER, S. D., P. A. LUCE, and D. B. PISONI. 1989. Priming Lexical Neighbors of Spoken Words: Effects of Competition and Inhibition. *JML*, 28.501–18.

GOLDINGER, S. D., T. AZUMA, H. M. KLEIDER, and V. M. HOLMES. 2003. Font-Specific Memory: More than Meets the Eye? In Bowers and Marsolek 2003, 157-96.

GORDON, B. 1983. Lexical Access and Lexical Decision: Mechanisms of Frequency Sensitivity. *JVLVB*, 22.24-44.

GROSSBERG, S. 1995. Neural dynamics of motion perception, recognition learning, and spatial attention. In *Mind as Motion*, ed. by R. Port and T. vanGelder, 449–90. Cambridge, MA: MIT Press.

HALL, G. 2003. Learned Changes in the Sensitivity of Stimulus Representations: Associative and Nonassociative mechanisms. *QJEP*, 56B.43-55.

HAY, J. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics* 39.1041-70.

HINTZMAN, D. L. 1986. 'Schema abstraction' in a multiple-trace memory model. *PR*, 105.251-79.

HOOPER, J. B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In *Current Progress in Historical Linguistics*, ed. by W. Christie. Amsterdam: North Holland.

HUNTSMAN, L. A., and S. D. LIMA. 2002. Orthographic Neighbors and Visual Word Recognition. *JPR*, 31.289-306.

JESCHENIAK, J. D., and W. J. M. LEVELT. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *JEP: LMC* 20.824-43.

KAMIN, L. J. 1969. Predictability, Surprise, Attention, and Conditioning. In *Punishment and Aversive Behavior*, ed. by B.Campbell, and R.Church, 279–98. New York: Appleton-Century-Crofts.

KAPATSINSKI, V. M. 2006. Having something common in common is not the same as sharing something special: Evidence from sound similarity judgments. Paper presented at LSA Annual Meeting, Albuquerque, NM. Available at http://www.unm.edu/~alator

KAPATSINSKI, V. M. 2004. Phonological similarity relations in the mental lexicon: Network organization of the lexicon and phonology. Paper presented at VIII Encuentro Internacional de Linguistica en el Noroeste, Hermosillo. Available at http://www.unm.edu/~alator

KORIAT, A. 1981. Semantic Facilitation in Lexical Decision as a Function of Prime-Target Association. *Memory and Cognition*, 9.587-598.

LUCE, P. A., S. D. GOLDINGER, E. T. AUER, and M. S. VITEVITCH. 2000. Phonetic Priming, Neighborhood Activation, and PARSYN. *PP*, 62.615–25.

LUCE, P. A., and D. B. PISONI. 1998. Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and Hearing*, 19.1– 36.

MARSLEN-WILSON, W., L. TYLER, R. WAKSLER, and L. OLDER. 1994. Morphology and meaning in the English mental lexicon. *PR*, 101, 3-33.

MCCLELLAND, J. L., B. L. MCNAUGHTON, and R. C. O'REILLY. 1995. Why There are Complementary Learning Systems in the Hippocampus and Neocortex: Insights from the Successes and Failures of Connectionist Models of Learning and Memory. *PR*, 102, 419–57.

MCCLELLAND, J. L. and D. E. RUMELHART. 1981. An Interactive Activation Model of Context Effects in Letter Perception: Part I. An Account of Basic Findings. *PR*, 88.375-407.

MCGOVERN, J. B. 1964. Extinction of Associations in Four Transfer Paradigms. *Psychological Monographs*, 78.

MCLENNAN, C., P. A. LUCE, and J. CHARLES-LUCE. 2003. Representation of Lexical Form in Spoken Word Recognition. *JEP: LMC*, 29.539-53.

METSALA, J. L. 1997. An Examination of Word Frequency and Neighborhood Density in the Development of Spoken-Word Recognition. *Memory and Cognition*, 25.47-56.

MILLER, J. L. 1994. On the Internal Structure of Phonetic Categories: A Progress Report. *Cognition*, 50.271-85.

MODER, C. L. 1992. *Productivity and Categorization in Morphological Classes*. Ph.D. Dissertation: SUNY Buffalo.

MOSCOSO DEL PRADO MARTIN, F., M. ERNESTUS, and R. H. BAAYEN. 2004. Do Type and Token Frequency Effects Reflect Different Mechanisms? *Brain and Language*, 90.287-98.

MURRAY, W. S., and K. I. FORSTER. 2004. Serial Mechanisms in Lexical Access: The Rank Hypothesis. *PR*, 111.721-56.

NATION, K., and M. J. SNOWLING. 1998. Individual Differences in Contextual Facilitation: Evidence from Dyslexia and Poor Reading Comprehension. *Child Development*, 69.996-1011.

NEELY, J. H. 1991. Semantic Priming Effects in Visual Word Recognition: A Selective Review of Current Findings and Theories. In *Basic processes in reading: Visual word recognition*, ed. by D.Besner and G.W.Humphreys. Hillsdale: Erlbaum.

NEWMAN, R. S., and D. J. GERMAN. 2002. Effects of Lexical Factors on Lexical Access among Typical Language-Learning Children and Children with Word-Finding Difficulties. *Language and Speech*, 45.285-317.

NORRIS, D. G. 1984. The Effects of Frequency, Repetition and Stimulus Quality in Visual Word Recognition. *QJEP*, 36A.507-18.

PAGLIUCA, W., and R. MOWREY. 1987. Articulatory Evolution. In *Papers from the 7th Intenrational Conference on Historical Linguistics*, ed. by A. G. Ramat, O. Carruba, and G.

Bernini. Amsterdam: Benjamins.

PEREA, M., and E. ROSA. 2000. Repetition and Form Priming Interact with Neighborhood Density at a Short Stimulus-Onset Asynchrony. *Psychonomic Bulletin and Review, 7*.668-77.

PERFETTI, C. A., and T. HOGABOAM. 1975. Relationship between Single-Word Decoding and Reading Comprehension Skill. *Journal of Educational Psychology*, 67.461-9.

PIERREHUMBERT, J. 2001. Exemplar Dynamics: Word Frequency, Lenition and Contrast. In F*requency and the Emergence of Linguistic Structure*, ed. by J. Bybee and P. Hopper. Amsterdam: Benjamins.

PLAUT, D. C., and J. R. BOOTH. 2000. Individual and Developmental Differences in Semantic Priming: Empirical and Computational Support for a Single-Mechanism Account of Lexical Processing. *PR*, 107.786-823.

PULVERMULLER, F. 2003. *The Neuroscience of Language: On Brain Circuits of Words and Serial Order*. Cambridge, UK: CUP.

RADEAU, M., J. MORAIS, and J. SEGUI. 1995. Phonological priming between monosyllabic spoken words. *JEP: HPP*, 21.1297-1311.

RATCLIFF, R., and G. MCKOON. 1981. Does Activation Really Spread? *PR*, 88.454-62.

RATCLIFF, R., and G. MCKOON. 1988. A retrieval theory of priming in memory. *PR*, 95.383-408.

RUECKL, J. G. 2003. A Connectionist Perspective on Repetition Priming. In Bowers and Marsolek (2003), 67-104.

RYLE, G. 1949. *The Concept of Mind*. London: Hutchinson.

SCARBOROUGH, D. L., C. CORTESE, and H. S. SCARBOROUGH. 1977. Frequency and Repetition Effects in Lexical Memory. *JEP: HPP*, 3.1-17.

SCHRIEFERS, H., A. FREDERICI, and P. GRAETZ. 1992. Inflectional and Derivational Morphology in the Mental Lexicon: Symmetries and Asymmetries in Repetition Priming. *QJEP*, 44A.373-90.

SCHUBERTH, R. E., and P. EIMAS. 1977. Effects of Context on the Classification of Words and Non-Words. *JEP: HPP*, 3.27-36.

SIMPSON, G. B., and T. C. LORSBACH. 1983. The Development of Automatic and Conscious Component of Contextual Facilitation. *Child Development*, 54.760-72.

SOLAN, Z., D. HORN., E. RUPPIN, and S. EDELMAN. 2005. Unsupervised learning of natural languages. *PNAS*, 102.11629-34.

STARK, C. E. L. 1997. *Repetition Priming of Words, Pseudowords, and Nonwords.* Ph.D. Dissertation: Carnegie Mellon University.

STEYVERS, M., and J. B. TENENBAUM. 2005. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive Science* 29.41-78.

STOCKALL, L. 2004. *Magnetoencephalographic Investigations of Morphological Identity and Irregularity*. Ph.D. Dissertation, MIT.

SUMIDA, R. A., and M. G. DYER. 1992. Propagation Filters in PDS Networks for Sequencing and Ambiguity Resolution. *Neural Information Processing Systems*, 4.233-40.

SUMIDA, R. A. 1997. *Parallel Distributed Semantic Networks for Natural Language Processing*. Ph.D. Dissertation, UCLA.

TENPENNY, P. L. 1995. Abstractionist versus Episodic Theories of Repetition Priming and Word Identification. *Psychonomic Bulletin and Review*, 2.339-63.

THOMSEN, C. J., H. LAVINE, and J. KOUNIOS. 1996. Social Value and Attitude Concepts in Semantic Memory: Relational Structure, Concept Strength, and the Fan Effect. *Social Cognition*, 14.191-225.

VERSACE, R., and B. NEVERS. 2003. Word Frequency Effect on Repetition Priming as a Function of Prime Duration and Delay between Prime and Target. *British Journal of Psychology*, 94.389-408.

VITEVITCH, M. S., and P. A. LUCE. 1998. When Words Compete: Levels of Processing in Perception of Spoken Words. *Psychological Science*, 9.325-9.

WHITTLESEA, B. W. A. 2003. On the Construction of Behavior and Personal Experience: The Production and Evaluation of Performance. In Bowers and Marsolek (2003), 239-60.

ZEELENBERG, R. 1998. *Testing Theories of Priming*. Ph. D. Dissertation: U of Amsterdam.

ZEVIN, J. D., and M. S. SEIDENBERG. 2004. Age-of-Acquisition Effects in Reading Aloud: Tests of Cumulative Frequency and Frequency Trajectory. *Memory and Cognition*, 32.31-