

**PRODUCTIVITY OF RUSSIAN STEM EXTENSIONS:  
EVIDENCE FOR AND A FORMALIZATION OF NETWORK THEORY**

**BY**

**VSEVOLOD M. KAPATSINSKI**

**A.A., LIBERAL ARTS, SIMON'S ROCK COLLEGE OF BARD, 2002  
B.A., LINGUISTICS, UNIVERSITY OF NEW MEXICO, 2003**

THESIS

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Master of Arts  
Linguistics**

The University of New Mexico  
Albuquerque, New Mexico

**2005**

© 2005, Vsevolod M. Kapatsinski

## **DEDICATION**

**To my mother, Varvara Kapatsinskaia.**

## ACKNOWLEDGEMENTS

Many thanks to Joan Bybee, Jill Morford, Caroline Smith, Rena Torres-Cacoulllos, and Catherine Travis for insightful comments on and stimulating discussions of earlier drafts of this paper.

My thanks also go to Farrell Ackerman, John Alderete, Clay Beckner, Lev Blumenfeld, Eve Clark, Caitlyn Dillon, Paul Kiparsky, James McClelland, James Macfarlane, Ana A. Medina-Murillo, Janet Pierrehumbert, David Pisoni, Robert Port, Adam Werle, the audiences at FASL 13, 4<sup>th</sup> International Conference on the Mental Lexicon, HDLS VI, VIII Encuentro Internacional de Linguística en el Noroeste, BLS 31, and CLS 41, and the anonymous reviewers of the FASL 13 proceedings, for interesting questions, comments, and suggestions for improvement.

I would like to thank Harald Clahsen for giving me a firm grounding in the Dual Mechanism Theory and getting me interested in testing its predictions, Joan Bybee for introducing me to Network Theory and many interesting discussions of the issues described in the thesis, and to Jay McClelland for teaching me how to build connectionist models and discussions of the finer details of how the models work.

Finally, many thanks to Alec Marantz for presenting the poster at the 4<sup>th</sup> International Conference on the Mental Lexicon.

Of course, all mistakes of reasoning, methods, and analysis, slips of the hand and those of the mind, from the understandable to the unforgivable that may remain in this thesis are mine and mine alone.

**PRODUCTIVITY OF RUSSIAN STEM EXTENSIONS:  
EVIDENCE FOR AND A FORMALIZATION OF NETWORK THEORY**

**BY**

**VSEVOLOD M. KAPATSINSKI**

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the  
Requirements for the Degree of

**Master of Arts  
Linguistics**

The University of New Mexico  
Albuquerque, New Mexico

**April, 2005**

# **EVIDENCE FOR AND A FORMALIZATION OF NETWORK THEORY**

**by**

**Vsevolod M. Kapatsinski**

A.A., Liberal Arts, Simon's Rock College of Bard, 2002

B.A., Linguistics, University of New Mexico, 2003

## **Thesis committee:**

Joan Bybee (director)

Jill Morford

Caroline Smith

## **ABSTRACT**

This thesis presents experimental and corpus-based evidence that 1) contrary to the Dual Mechanism Model (Pinker and Prince 1994) there are morphological domains without a rule-based default, 2) both types and tokens are stored, 3) sublexical phonological units can exhibit partial overlap and get associated with co-occurring morphemes, 4) contrary to Albright and Hayes (2003), such associations can form at a distance, and 5) naturalness is determined based on the properties of the presented stimulus and not based on a comparison of the presented stimulus to possible alternatives. A novel version of Network Theory (Bybee 1988, 2001) formalized as a localist associative network with type and token nodes and spreading activation is argued to be most consistent with evidence from morphological productivity, priming, word recognition, and associative learning. Ways to test the proposed theory against connectionist models are discussed.

## TABLE OF CONTENTS

<b>LIST OF FIGURES</b> .....	xii
<b>LIST OF TABLES</b> .....	xiii
<b>CHAPTER 1: INTRODUCTION</b> .....	1
<b>CHAPTER 2: MODELS OF MORPHOLOGICAL PRODUCTIVITY</b> .....	6
<b>2.1. Connectionist networks</b> .....	6
2.1.1. Basic features.....	6
2.1.2. Successes .....	7
2.1.3. Criticisms.....	11
<b>2.2. The Dual Mechanism Model</b> .....	21
2.2.1. Basic features.....	21
2.2.2. Successes .....	22
2.2.3. Criticisms.....	23
<b>2.3. Rule-Based Learner</b> .....	31
2.3.1. Basic features.....	31
2.3.2. Successes .....	32
2.3.3. Criticisms.....	34
<b>2.4. Network theory</b> .....	40
2.4.1. Basic features.....	40
2.4.2. Successes .....	41
2.4.3. Criticisms.....	46
<b>2.5. Local Activation Spread Theory</b> .....	49
2.5.1. Basic features.....	50

2.5.1.1. Architecture.....	50
2.5.1.2. Microstructure.....	51
2.5.1.3. Dynamics of activation spread.....	52
2.5.1.4. Empirical constraints on link structure.....	53
2.5.2. The LAST account of the findings.....	55
<b>2.6. Conclusion</b> .....	66
<b>CHAPTER 3: CONTROLLING FOR SOUND SIMILARITY</b> .....	68
<b>3.1. The FIRM method of controlling for similarity</b> .....	68
<b>3.2. Empirical findings on sound similarity</b> .....	70
3.2.1. Previous work.....	70
3.2.1.1. Major issues and methods.....	70
3.2.1.2. Location of mismatch in relation to word boundaries.....	73
3.2.1.3. Length of match and mismatch.....	78
3.2.1.4. Psychological reality of sublexical phonological units.....	84
3.2.1.5. Stress.....	89
3.2.1.6. Type of mismatch.....	89
3.2.1.7. Summary.....	90
3.2.2. Experiment A: Number of mismatched syllabic constituents, location of mismatch and number of matched segments.....	91
3.2.2.1. Method.....	92
3.2.2.2. Number of mismatched syllabic constituents and mismatch location.....	96
3.2.2.3. Number of matched segments.....	98

3.2.3. Experiment B: Segments and milliseconds .....	99
3.2.3.1. Method .....	99
3.2.3.2. Results and discussion.....	100
3.2.4. General discussion: Implications for the study of morphological productivity.....	101
<b>3.3. <i>The similarity metrics used in this study</i></b> .....	102
<b>3.4. <i>Future directions</i></b> .....	105
<b>CHAPTER 4: EXPERIMENT I: DEFINING ATTRIBUTES OF THE RULE-     BASED DEFAULT ARE DISSOCIABLE</b> .....	108
<b>4.1. <i>Introduction</i></b> .....	108
<b>4.2. <i>Methods</i></b> .....	109
4.2.1. The task and the participants .....	109
4.2.2. Stimuli .....	110
4.2.2.1. Measuring sound similarity.....	110
4.2.2.2. Controlling for priming .....	111
4.2.2.3. Root-final consonant .....	112
<b>4.3. <i>Results</i></b> .....	113
4.3.1. Palatalization of the root-final consonant.....	113
4.3.2. Similarity effects and productivity .....	114
<b>4.4. <i>Discussion</i></b> .....	116
<b>4.5. <i>Appendix: Results in detail</i></b> .....	117
<b>CHAPTER 5: CORPUS STUDY I: EVENNESS OF DISTRIBUTION AND     SENSITIVITY TO SIMILARITY</b> .....	119

5.1. <i>Introduction</i> .....	119
5.2. <i>Methods</i> .....	119
5.3. <i>Results and discussion</i> .....	120
<b>CHAPTER 6: EXPERIMENT II: ALL THE POPULAR SUFFIXES ARE SENSITIVE</b> .....	131
6.1. <i>Introduction</i> .....	131
6.2. <i>Methods</i> .....	132
6.3. <i>Results and discussion</i> .....	135
<b>CHAPTER 7: EXPERIMENT III: RIMES AND BODIES MATTER</b> .....	143
7.1. <i>Introduction</i> .....	143
7.2. <i>Methods</i> .....	143
7.3. <i>Results and discussion</i> .....	146
<b>CHAPTER 8: CONCLUSION</b> .....	151
8.1. <i>Summary</i> .....	151
8.2. <i>Future directions</i> .....	153
8.2.1. Source-oriented vs. product-oriented schemas: Preliminary results .....	153
8.2.2. Schemas vs. analogy .....	156
8.2.3. Similarity metrics .....	157
8.2.4. Transitional probability versus constituent frequency .....	158
8.2.5. LAST vs. connectionist models .....	158
8.3. <i>Conclusion</i> .....	161
<b>REFERENCES</b> .....	162

## LIST OF FIGURES

Figure 2.1. The activation-functional account of frequency effects and individual differences in priming .....	11
Figure 2.2. Architecture of memory according to LAST .....	51
Figure 2.3. A priori possible LAST link structures .....	52
Figure 2.4. The effect of the frequency of the prime on amount of priming hypothesized by LAST .....	59
Figure 5.1. Type frequencies with which different velar consonants occur in the root-final position in verbs containing -a- .....	122
Figure 5.2. Type frequencies with which different sounds occur in the penultimate position in verbs containing -a- that have a velar root-final consonant. ....	122
Figure 5.3. Type frequencies with which different sounds occur in the antepenultimate position in verbs containing -a- that have a velar root-final consonant. ....	123
Figure 5.4. Type frequencies with which different labial consonants occur in the root-final position in verbs containing -a- .....	124
Figure 5.5. Type frequencies with which different labial consonants occur in the root-final position in verbs containing -i- .....	124
Figure 5.6. Type frequencies with which different phonemes occur in the penultimate position in verbs containing -a- that have a labial root-final consonant. ....	125
Figure 5.7. Type frequencies with which different phonemes occur in the penultimate position in verbs containing -i- that have a labial root-final consonant. ....	125
Figure 5.8. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -a- that have a labial root-final consonant. ....	126
Figure 5.9. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -i- that have a labial root-final consonant. ....	126
Figure 5.10. Type frequencies with which different coronal and palatal consonants occur in the root-final position in verbs containing -a- .....	127
Figure 5.11. Type frequencies with which different coronal and palatal consonants occur in the root-final position in verbs containing -i- .....	127
Figure 5.12. Type frequencies with which different phonemes occur in the penultimate position in verbs containing -a- that have a coronal or palatal root-final consonant. ....	128
Figure 5.13. Type frequencies with which different phonemes occur in the penultimate position in verbs containing -i- that have a coronal or palatal root-final consonant. ....	128
Figure 5.14. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -a- that have a coronal or palatal root-final consonant. ....	129
Figure 5.15. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -i- that have a coronal or palatal root-final consonant. ....	129

## LIST OF TABLES

Table 3.1. Matched and mismatched phonemes and their locations in stimuli classified as having high or low phonological overlap in Stockall (2004).....	82
Table 3.2. Stimuli used in experiment A .....	94
Table 3.3. Mismatch salience as a function of location in bisyllabic words .....	96
Table 3.4. Mismatch in two contiguous segments in a single constituent that include the word-final segment versus mismatch in two medial contiguous segments in two constituents.....	97
Table 3.5. Mismatch in two contiguous segments in one constituent vs. two non-contiguous segments in two constituents where one of the segments is word-final in both cases.....	97
Table 3.6. Mismatch in two contiguous segments in two constituents vs. two non-contiguous segments in two constituents where none of the segments are word-final in both cases.....	97
Table 3.7. Mismatch in two non-contiguous segments in two syllables vs. in one syllable .....	98
Table 3.8. Mismatches beginning in a word-medial consonant close to word beginning vs. those beginning in a medial consonant close to word end.....	98
Table 3.9. Effect of number of matched segments .....	99
Table 3.10. Stimuli used in experiment B.....	100
Table 3.11. Words differing by one segment vs. two contiguous segments: No differences in temporal duration of mismatched stretches of speech.....	100
Table 3.12. Words differing by one segment vs. two contiguous segments: Increased differences in temporal duration of mismatched stretches of speech	101
Table 4.1. Stimuli.....	113
Table 4.2. Palatalization.....	114
Table 4.3. Palatalized root-final consonants by stimulus similarity .....	114
Table 4.4. Affix productivity after velars. X stands for ‘the class of roots that take the stem extension shown on the right’ .....	115
Table 4.5. Affix productivity after labials. ....	116
Table 4.6. Affix productivity after coronals and palatals .....	116
Table 4.7. Velar-final roots.....	117
Table 4.8. Labial-final roots.....	117
Table 4.9. Coronal- and palatal-final roots.....	118
Table 5.1. Slopes of lines in figures 5.1-5.16, where the smaller the absolute value of the slope, the more evenly the words containing the stem extension are distributed in phonological space .....	121
Table 6.1. Stimuli used to test the size of similarity effects or -a- and -i-.....	133
Table 6.2. Stimuli used to test the influence of the root-final consonant .....	134
Table 6.3. Acceptability of nonce verbs bearing -a- and -i- depending on the contents of their phonological neighborhoods.....	136
Table 6.4. Distance to nearest neighbor and sensitivity to similarity .....	137
Table 6.5. Naturalness ratings of (coronal- and palatal-final) verbs derived from nouns similar to no existing words with -a-, -i-, or -ova-.....	138
Table 6.6. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /b/ vs. /d/ .....	141

Table 6.7. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /p/ vs. /t/ .....	141
Table 6.8. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /m/ vs. /n/ .....	141
Table 6.9. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /v/ vs. /z/ vs. /ʒ/.....	142
Table 6.10. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva- .....	142
Table 7.1. Stimuli used for testing whether rimes and bodies are associated with stem extensions they co-occur with .....	146
Table 7.2. Body influence on naturalness judgments .....	147
Table 7.3. Rime influence on naturalness judgments .....	148

## CHAPTER 1: INTRODUCTION

The mechanisms that determine the relative productivity of synonymous affixes have become the subject of vigorous debate since Rumelhart and McClelland (1986) published their model of English past tense formation that provided an alternative to the traditional idea that all regularity had to be stated in terms of maximally general rules with the lexicon being a “prison for the lawless” (DiSciullo and Williams 1987). As Bybee and Slobin (1982) and Bybee and Moder (1983) pointed out, English verbs with irregular past tenses are not distributed evenly throughout the lexicon but form tight clusters which are partially productive being able to attract new members and the patterns they share are even overgeneralized by children acquiring English, although to a lesser extent than is the regular past tense.

Rumelhart and McClelland (1986) were able to model the U-shaped acquisition curve in the English past tense with a connectionist network that did not use any symbolic categorical rules and concluded that such rules were at best imperfect approximations of actual human behavior. Pinker and Prince (1988) criticized the model and argued that while productivity of irregular patterns depends on the strength of those patterns within the associative network, regular inflection is a symbolic rule that applies in default circumstances when the speaker fails to form an irregular by analogy to existing irregulars in time.

The two findings used by proponents of Pinker and Prince’s Dual Mechanism Model to justify the regular/irregular distinction as psychologically real is that 1) there are cases where the default morph is not the most frequent morph in the set of competitors in terms of its type frequency (McCarthy and Prince 1990, Clahsen et al.

1992, Clahsen 1999, Berent et al. 1999), and 2) the default morph is (by definition) the most frequently used morph with nonce words that are similar to no existing words and it is applied to such words as or more frequently than it is applied to nonce words similar to words to which the morph attaches in the lexicon (Clahsen 1999, Pinker and Prince 1988, 1994; Pinker 1999, Berent et al. 1999), i.e. it is subject to neither type frequency nor similarity effects. In chapter 4, I demonstrate that the defining attributes of the rule-based default listed in (2) are dissociable. In chapter 5, I show that difficulty with finding similarity effects correlates with evenness of distribution of the words containing the morpheme in phonological space. Having introduced a plausible metric of phonological similarity to use in the study in chapter 3, I show that this is only a methodological difficulty and that when controls for the actual difference in similarity are in place different competitor morphemes are equally sensitive to it or at least very similar in sensitivity in chapter 6. Constraints on possible metrics are discussed in chapter 3.

Albright and Hayes (2003) found that a similarity effect is observed for the regular English past tense where nonce verbs that are similar to existing regulars are more likely to take the regular than nonce verbs similar to neither regulars nor irregulars and when subjects are asked to rate the goodness of regular and irregular past tenses of the verb provided by the experimenter, they favor the regular version more when the nonce verb is similar to many regulars than when it is similar to neither regulars nor irregulars. Based on this finding, Albright and Hayes argue that rules are equipped with context-specific indices of reliability such that the more reliable the rule is in a given context and the more types it applies to in the lexicon as a whole, the more likely it is to apply. In Albright and Hayes' model, all patterns are applied via stochastic rules whose application

probability varies depending on their reliability and no past tenses are formed by analogy to existing lexical items.

Finally, Network Theory (Bybee 1985, 1988, 1995, 2001) proposes that the mind is a network in which every word corresponds to a node or set of nodes. The sets of nodes corresponding to distinct word types do not overlap but similarity strengthens connections between words. New words forms are formed by analogy to existing word forms of the same kind. The more types instantiate a pattern, the more likely it is that the pattern would be applied to a nonce word similar to no existing words. Network theory points out that the likelihood of using a morph with a given nonce word often depends more on the number of words containing that morph that are in the word's neighborhood (a product-oriented factor) than on the frequency of the course-oriented rule adding the morph to stems similar to the nonce word (Wang and Derwing 1994). In addition, many inflectional domains lack a clear default. For instance, Dabrowska (2001) found that no morph was a default for Polish genitive singular and By adopting a localist<sup>1</sup> representation and Exemplar Theory (Hintzman 1986, Nosofsky 1988, Pierrehumbert 2001) which maintains that a separate representation is formed for every token of a word, Network Theory is able to handle sequential structure (Marcus 1998), dissociations between token and type frequency effects (Bybee 2001), as well as between direct and associative activation (Hall 2003) which connectionist networks have trouble handling. Dissociations between type frequency and productivity are handled by appealing to confounding factors.

---

<sup>1</sup> Localist representations are representations in which every unit is represented as a separate node or a set of nodes that are not used for encoding any other unit. It is the opposite of distributed representation, in which units are represented as overlapping patterns of activation where a single node can be used to represent several units.

Network Theory differs from Albright and Hayes' Rule-Based Learner in its similarity metric: Albright and Hayes believe that when the nonce word is compared to existing words the speaker starts from phonemes adjacent to the morpheme boundary at which morph choice is to be made and works outward. If s/he encounters a mismatch, further phonemes are not compared. By contrast, in Network Theory, all phonemes are taken into account, although mismatches are assigned different weights depending on their position in the word, including such factors as whether it is adjacent to a word boundary (derived from primacy and recency effects) and whether it comprises an entire syllabic constituent or consonant cluster. The Network Theory metric has the advantage of being domain-general rather than specific to morphological processing. However, since current data come from phonological processing of nonce words, it is important to show that the domain-general similarity metric can account for morphological productivity. In chapter 7, I show that the similarity metric proposed by Albright and Hayes (2002, 2003) fails to account for the effect of root beginning-based schemas on suffix productivity and that structured similarity is in fact predicted to be stronger than variegated similarity by LAST (introduced in section 2.5), an exemplar-based localist formalization of Network-theoretic principles in a spreading-activation framework (Collins and Loftus 1975, Anderson 1976, 1983, 2000) introduced in chapter 2, where a more thorough review of the existing models of the lexicon is conducted.

In chapter 8, I conclude that 1) contrary to the Dual Mechanism Model (Pinker and Prince 1994) there are morphological domains without a rule-based default, 2) both types and tokens are stored, 3) sublexical phonological units can exhibit partial overlap and get associated with co-occurring morphemes, 4) contrary to Albright and Hayes

(2003), such associations can form at a distance, and 5) naturalness is determined based on the properties of the presented stimulus and not based on a comparison of the presented stimulus to possible alternatives. I discuss future directions, including ways to test whether the mind uses distributed or localist representation, whether schemas or whole-form analogy best account for productivity, whether morphological decomposition is obligatory, and whether the generalizations involved are product-oriented or source-oriented.

## **CHAPTER 2: MODELS OF MORPHOLOGICAL PRODUCTIVITY**

### ***2.1. Connectionist networks***

#### **2.1.1. Basic features**

A connectionist network must include at least two sets of nodes: the input layer and the output layer. Different words are represented as different patterns of activation over the same set of nodes. During training, the network is presented with both the stem form encoded over the input layer and the correctly inflected form over the output nodes. In all networks explored so far, the network first sees the input and is allowed to produce any output it wishes. The weights of the connections between input and output nodes are then adjusted based on how much the connection contributed to the network's deviation from the correct output. The output of a connection is a function of the input and the weight such that the greater the weight, the more of the input activation reaches the output node. If the connection contributes greatly to the network's error its weight is decreased so it contributes less to the network's output in the future. Typically, every node in one layer is connected to every node in the other layer and all connections are trainable. At testing, the network is presented with inputs it has not encountered before and its output is measured and compared with the productions of human subjects. As training progresses, the network's performance on the words in the training set asymptotically approaches perfection but the network's generalization performance (performance over the testing set) begins to decrease after a certain point (the network becomes overtrained). The training is typically stopped and the testing begins as late as possible but before generalization performance starts to decrease. The network's nodes typically have

sigmoid activation functions so that as its activation value approaches upper and lower limits, the same increases in input activation lead to smaller increases in output activation levels, hence the nodes tend to settle into binary response patterns. The greater the frequency of an input-output pairing, the better it will be learned, although the same difference in presentation frequency will have a greater effect if the frequencies of the compared pairings are low rather than high because of the sigmoid activation function. Both lower and upper limits correspond to high frequency types because words are represented as patterns of +1's and -1's and the node activation levels in a well-learned (frequently presented and/or low-neighborhood-density) representation will have values approaching the correct pattern of +1's and -1's while a poorly learned representation is likely to be instantiated as a relatively uniform pattern across nodes with activation levels clustering around 0.

### **2.1.2. Successes**

Perhaps, the most important result of the Rumelhart and McClelland (1986) simulation is that the network exhibited the classic U-shaped acquisition curve where the child starts off producing the irregulars correctly, then overgeneralizes the regular pattern and erroneously applies it to (some) irregular verbs (some of the time) and finally gradually re-acquires the irregulars (Berko 1958, Brown 1973, Cazden 1968, Ervin 1964, Kuczaj 1977, 1981, Marcus et al. 1992). The onset of overgeneralization has been claimed to be the point of acquisition of the categorical rule but the curve turns out to be observable even if no rules are used.

Another success of the McClelland and Rumelhart (1986) model is that it was

able to replicate the finding that children tend to use the stem as the past tense form of verbs that already end in /t/ or /d/ (Stemberger 1981, Bybee and Slobin 1982). A related finding is that past tense patterns consisting of a vowel change combined with a coronal stop suffix are learned easier by the model than past tense patterns that consist of just a single vowel change. This is due to a product-oriented generalization (Bybee 1995, 2001) in that nodes corresponding to word-final coronal stops in the past tense (output) layer had a high resting activation level due to high frequency of such stops in the past tense.

In addition, the model predicted that irregular classes that require changes to the stem are harder to learn than classes that do not require such changes. Supporting evidence is provided by Dabrowska (2001) who found that the same children overgeneralize a single Polish genitive plural suffix while none of the genitive singular suffixes are overgeneralized more than others, although both domains feature an affix that was of higher type frequency than its competitors. Interestingly, the most common plural affix competes only with affixes that require stem changes while none of the singular affixes require stem changes.

Finally, the model was able to generalize to verbs it had not encountered during training with 90% accuracy, which is the standard criterion for having acquired a grammatical pattern for humans (e.g. Van Kampen 1997).

Plaut and Booth (2000) have applied a connectionist model to account for frequency effects and individual differences in priming and were successful at reproducing a complex pattern of results. They proposed that the result that high frequency primes and targets exhibit less priming can be accounted for by sigmoid activation functions of the nodes in a connectionist model of associative memory. Just as

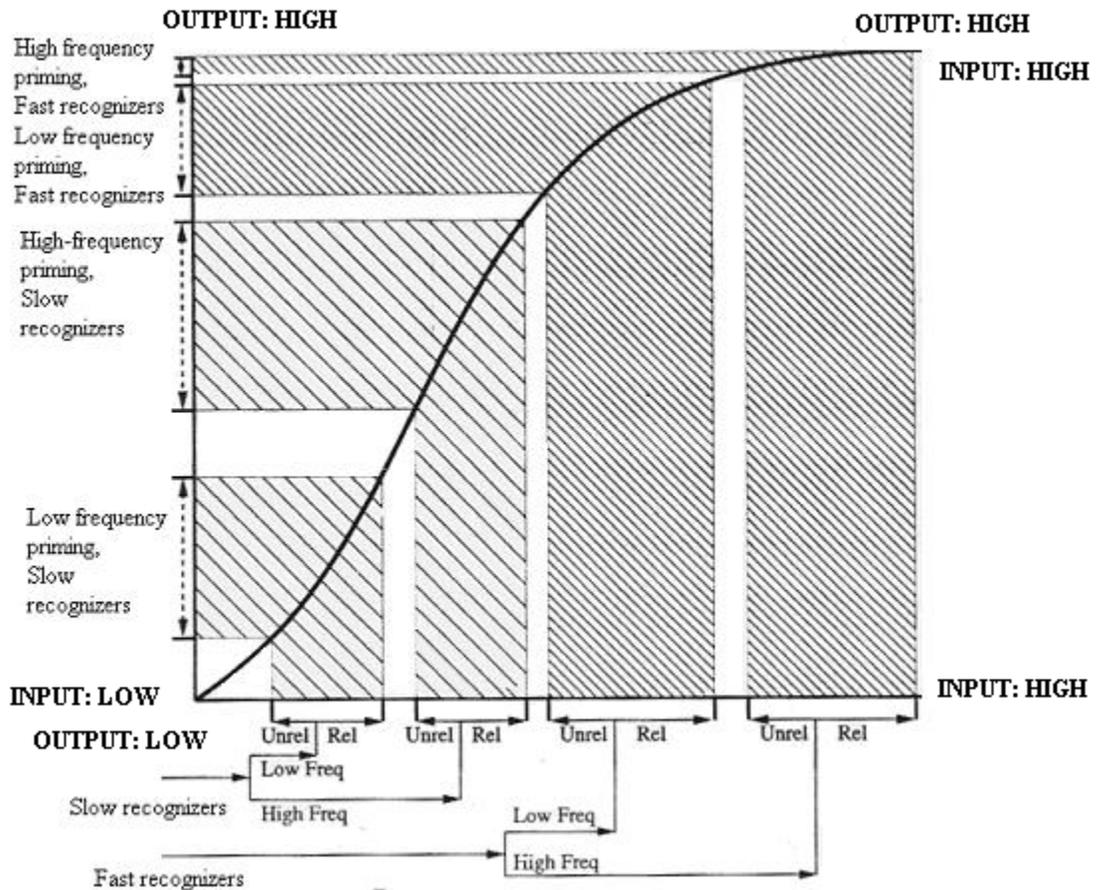
in Rumelhart and McClelland's (1986) model, words are superimposed on a fixed set of nodes encoding their phonological and semantic characteristics. The activation function specifies how much output activation is produced given the amount of input activation the node receives. By incorporating the assumption that the same increase in input activation will produce a smaller increase in output activation if the resting activation level is high into the activation function, Plaut and Booth are able to predict the finding that high-frequency words exhibit less priming than low-frequency words (cf. Scarborough et al. 1977, Jacoby and Dallas 1981, Jacoby 1983, Forster and Davis 1984, Norris 1984, Jacoby and Hayman 1987, Stark 1997, Stark and McClelland 2000, Nevers and Versace 1998, Versace 1998, Perea and Rosa 2000, Versace and Nevers 2003 for identity/repetition priming, Moder 1992 for morphological priming, Schuberth and Eimas 1977, Schuberth et al. 1981, Becker 1979, Stanovich and West 1981, 1983, Stanovich et al. 1981, West and Stanovich 1982, Neely 1991, Borowsky and Besner 1993, Thomsen et al. 1996, and Plaut and Booth 2000 for semantic priming, Goldinger et al. 1989, Radeau et al. 1995, Luce et al. 2000 for phonological priming, and Perea and Rosa 2000 for orthographic priming).

In addition, because the words are superimposed activation patterns over a fixed set of nodes and because the nodes all have sigmoid activation functions, a word token to which the network is exposed early in its lifetime will have greater effect on the resting activation levels of the nodes than a word token experienced late in the networks lifetime (Bonin et al. 2001, Meschyan and Hernandez 2002, Morrison et al. 2002, Newman and German 2002, Zevin and Seidenberg 2002, 2004, Ghyselinck et al. 2004). Thus, the representations of early-acquired words will be more robust than those of late-acquired

words and will thus be accessed faster and by a wider range of inputs than representations of late-acquired words. Finally, because both age-of-acquisition and token frequency influence the same parameter, the resting activation level of the nodes in the network, early-acquired words will exhibit less priming, just as words with high token frequency (Barry et al. 2001).

In addition, Plaut and Booth (2000) found that the frequency effect on priming is larger for children who are fast at word recognition when their receptive vocabulary size is controlled. This is predicted by the sigmoid activation function by proposing that the lexical representations of children who are poor at word recognition receive less input activation. As shown in figure 2.1, if these children's input activation falls onto the linear portion of the curve whether or not the word is of high frequency, there will be no difference between the changes in output activation elicited by presentation of related high-frequency and low-frequency primes.

**Figure 2.1. The activation-functional account of frequency effects and individual differences in priming (Plaut and Booth 2000)**



### 2.1.3 Criticisms

Pinker and Prince (1988, 1994) have pointed out that the Rumelhart and McClelland (1986) model had no representation of morphological structure or semantics. Thus homophones would be treated as the same word while in reality they might exhibit different behavior in respect to the past tense in that denominal verbs are overwhelmingly regular even if a homophonous monomorphemic irregular exists (Kiparsky 1983, Williams 1981, Selkirk 1982, Kim et al. 1991).

However, there is nothing in the connectionist architecture per se that would exclude creating additional input units to encode semantic differences. In fact, examples

of models that do incorporate semantics exist, for instance, the Plaut et al. (1996) revision of the Seidenberg and McClelland (1989) reading model or the Joanisse and Seidenberg (1999) model of the English past tense formation. On the other hand, encoding morphemes into the input and output layers would mean using localist representations, at least for monomorphemic words, which is contrary to the most basic premise of connectionism. In addition, current connectionist models have no mechanism of adding new units to input and output layers based on how those units function in the input.

Ramscar (2002) systematically manipulated whether a nonce verb had a meaning that was similar to the meaning of a phonologically similar irregular verb or to that of a phonologically similar or dissimilar regular verb. When the nonce verbs ('frink' and 'sprink') were presented without context or in a context that indicated that they were semantically similar to 'drink', irregular inflection was preferred, while when the context indicated that the verbs were semantically similar to regulars 'wink', 'blink', or 'mediate', the regular was preferred. For some subjects, denominal status was established by presenting them with the verbs in a context in which a separate group of subjects had ranked the verbs as denominal after being presented with English denominal and non-derived verbs. No effect of denominal status was found. Thus it appears that adding semantic nodes, as has been done in several connectionist models, may be sufficient to account for the apparent effect of denominal status. However, Pinker and Ullman (2002) have criticized Ramscar's study by pointing out that only one verb was used.

Strong evidence for morphemes is provided by Stockall (2004), who pointed out that morphological priming produces more response facilitation and is more long-lived than semantic priming and that this cannot be accounted for by the presence of

phonological similarity in morphologically related words because phonological priming produces inhibition when word-beginnings, the location of the root in English verbs used in morphological priming studies, are primed (Goldinger et al. 1989, Radeau et al. 1995, Luce et al. 2000) and is even more short-lived than semantic priming. Furthermore, Stockall (2004) found, using event-related potentials on the magnetoencephalogram, that words that are both semantically and phonologically related but not morphologically related, e.g. ‘boil-broil’, ‘bloom-blossom’, ‘crinkle-wrinkle’, produce very little priming measured by the latency of the the M350 and behavioral reaction in the lexical decision task, while morphologically related primes speed up both M350 and reaction times.

These results can be accounted for by either a full decomposition or a dual-route model. In a full decomposition model, both the prime and the target are accessed through their component morphemes. In a dual-route model, the prime might be accessed directly and then activate its component morphemes. The target could then be accessed through the morphemes because activation of the prime will have pushed their resting activation level above the resting activation level of the target’s full-form representation, which might be inhibited by the prime (as assumed, for instance, by the interactive-activation model, Rumelhart and McClelland 1981). It is not clear whether the shared parts of semantically and phonologically related words are also stored but simply have very low resting activation levels as a result of very low type and token frequency or whether, as suggested by Stockall (2004), they are not stored at all. In either case, however, these results strongly support the notion of a morpheme.

Pinker and Prince (1988) point out that irregular plural forms are more acceptable within compounds than regular plural forms, e.g. *mice-eater* is more acceptable than *rats-*

*eater*. This finding is not restricted to affixes competing with stem changes because it is also found in German plural formation where *-s* is argued to be the default suffix and *-en* is its major competitor (Clahsen 1999). A factor that has not been given much attention until recently is the token frequency of the plural relative to the singular (Hay 2001). Hay (2001, 2003) suggested that affixes that derive forms that are more frequent than their stems are less productive than affixes deriving forms that are less frequent than their stems. It is well accepted that high token frequency of inflected irregular forms is what allows them to maintain their irregularity in the face of analogical pressure to regularize (Bybee 2001). Hay (2003) has demonstrated that derived forms with high relative frequency (of the derived relative to the stem) are more likely to serve as input to further derivation because the derived unit has a higher resting activation level than the stem and thus is likely to be accessed more easily when the base for further derivation is chosen. The same may be true of inflection: if irregular inflected forms tend to be of higher relative frequency, they may well be able to feed compounding better than regular inflected forms, which have lower relative frequency. Thus the difference between regulars and irregulars in the ability to feed compounding may be due to the inflected form's lexical strength (Bybee 1988, 1995, 2001) and not to whether the affix is attached via a symbolic rule or an analogical process. Hay (2003) assumes a dual-route lexical access process where a complex word form can be accessed either directly or through its component parts. However, relative frequency is also compatible with a full decomposition model, since a whole that is of low frequency relative to its parts has stronger competitor wholes that are fighting for the part's activation during lexical access than a whole with which the part is strongly associated, i.e. a whole with a high relative

frequency.

Pinker and Prince (1994) also criticize the Rumelhart and McClelland model for assuming that the child's input for the task of past tense acquisition largely consists of irregulars with the proportion of regulars increasing over time to explain the U-shaped acquisition curve. Marcus et al. (1992) found that the proportion of regulars to irregulars in the caregivers' speech remained constant throughout the child's acquisition of the past tense. Pinker and Prince (1988) found that the same is true of children's production. However, Storkel (2004) finds that age of acquisition of a word is positively correlated with its frequency when phonotactic probability, length, and neighborhood density are controlled, indicating that more frequent words are likely to be acquired first and thus the proportion of regulars to irregulars in the child's lexicon changes throughout development as hypothesized by Rumelhart and McClelland.

The model predicted double marking to occur for 7 verbs where the regular rule required the addition of *-t*, resulting in forms like /smouktəd/. Pinker and Prince (1988) claimed that this does not occur. While no systematic study of this issue has been conducted, this seems to be a plausible product-oriented generalization: word-final /təd/ does occur in many English past tenses. In my study of nonsense word perception (Experiment B in chapter 3), when 14 native English speakers were told to write down 150 pairs of nonsense words each, they wrote down a non-existent *-ed* at the end of a nonsense word that looked like a past tense of a verb in terms of containing a word-final alveolar stop following a consonant on 6 occasions, although no /əd/ was present in the input. The same finding has been obtained by Caitlyn Dillon (p.c.) in a population of cochlear implant users who were asked to repeat auditorily presented words. Double

marking was also observed in Koepcke's (1998) study of plural formation in English and German.

As Rumelhart and McClelland (1986) themselves point out, the model produced formations that were unlike their stems, namely 'membled' as the past tense of 'mail', 'toueder' as the past tense of 'tour', 'squakt' as the past tense of 'squat', 'maded' as the past tense of 'mate', and 'browned' as the past tense of 'brown'. Pinker and Prince (1988) pointed out that this does not occur with human subjects and that it is likely to be a consequence of distributed representation where units are not kept distinct. However, Bybee and Moder (1983) report that subjects do sometimes substitute final consonants of nonce verbs in English when forming the past tenses, e.g. 'smuk', 'stid' and 'stug' were all past tenses produced from 'smip'.

Two further problems stem from the fact that the model uses a fixed set of nodes to encode all words: 1) it cannot distinguish between the effects of type and token frequency as well as direct and associative activation and 2) it cannot store sequential structures with repeated units of the same type in an unambiguous manner in long-term memory.

Pinker and Prince (1988) first pointed out that connectionist models have trouble with unambiguous representation of sequences.

The problem can be illustrated with the following examples: if (1) is our long-term representation of /tɪk/, it is indistinguishable from /kɪt/: sequential order is lost if a word is simply a pattern of activation over a set of nodes with one node per phoneme of the language.

(1) [t ɪ k]

One solution that has been used (e.g. Plaut et al. 1996) is to fit all words into a single template in which phonemes activate different units depending on their positions in the word, e.g. ‘the first consonant in the coda of the first syllable’. The problem with this is that there is no fixed upper limit on word length thus the template would have to be infinite to accommodate any possible word.

Another solution has been to use wickelphonological encoding, that is, to use a single node for any sequence of three phonemes or features (Rumelhart and McClelland 1986). However, wickelphonology leads to unrealistic sound similarity judgments, e.g. ‘slit’ and ‘silt’ are judged by the model to be completely unlike each other because they share no wickelphones (Pinker 1999). Furthermore, no phonological rules apply to wickelphones (Pinker and Prince 1988). Finally, wickelphones do not provide a principled solution to the problem, since they can be repeated in a single word and the model has no way of representing two individuals as opposed to one strongly activated individual (Marcus 1998). For instance, ‘algal’ and ‘algalgal’ are different words in Oy kangand, an Australian language (Pinker 1999).

Even if sequentiality were encoded, the model would be still unable to unambiguously represent ‘hippopotamus’ (2): the representation is ambiguous between ‘hipppopotamus’, ‘hipoppotamus’, ‘hippopotamus’, etc. because the representation includes no information on how many ‘p’ and ‘o’ tokens are in the word as well as which tokens of ‘o’ are followed by ‘t’ and which by ‘p’.

(2) h → i → p → o → t → a → m → u → s

Lack of a distinction between tokens and types leads to the prediction that the higher the token frequency of a pattern, the more productive it should be since the

network was trained to apply the pattern more times. However, it is actually type frequency that makes a pattern more productive while token/type ratio reduces productivity. For instance, in French the first conjugation has the highest type frequency but the third conjugation has the highest token frequency and it is the first conjugation that is more productive (Guillaume 1927/1973, Bybee 2001). In the English past tense, vowels with the highest type frequency are the vowels that are used most frequently to form irregular past tenses and are not the same as the vowels that have the highest token frequency (Wang and Derwing 1994).

Many studies 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 (1994) found that speakers can reliably judge the goodness of different exemplars of a phoneme. More repetition priming is observed if the prime and target are said by the same voice than when they are produced in different voices (e.g. Craik and Kirsner 1974, Geiselman and Bellezza 1977, Goldinger et al. 1991, Goldinger 1992, 1996, Schachter and Church 1992, Palmeri et al. 1993, Church and Schachter 1994, see Lachs et al. 2003 and Goldinger et al. 2003 for reviews). Goldinger (1992) found that the effect of voice repetition persists for at least a week.

Even redundant features of the stimulus are stored, e.g. many studies found statistically significant font repetition effects such that words repeated in the same font produce more priming than words repeated in a different font (e.g. Hintzman and Summers 1973, Kirsner 1973, Jacoby and Hayman 1987, Blaxton 1989, Manso de Zuniga et al. 1991, Graf and Ryan 1990, Gibson et al. 1993, Goldinger et al. 2003).

Tenpenny (1995) conducted a meta-analysis on the studies that failed to find an effect of font repetition on the magnitude of repetition priming and found that all studies reported trends in the predicted direction and, taken together, the results reach statistical significance and provide further support for the hypothesis that redundant details are stored.

Section 3.2.3 of this thesis (see also Kapatsinski 2004e) shows 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, e.g. a [p<sup>h</sup>] that is 7 vs. 30 ms long.<sup>2</sup>

Moscato del Prado Martin et al. (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

---

<sup>2</sup> The words are reported correctly by subjects in both conditions as containing a /p/.

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 a connectionist model was trained over the same training corpus as the rule-based model, it found the 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.

Evidence for types comes from McLennan et al. (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 mapped onto the same representation, unlike primes and targets produced by different speakers (Goldinger 1992) or in different fonts (Tenpenny 1995, Goldinger et al. 2003), which produce less priming than primes and targets produced in the same voice or font.

Finally, evidence for the type/token distinction comes from a recent associative learning study. In associative learning, the more frequently a stimulus is presented, the weaker the response elicited by it (habituation effect). Furthermore, the more frequently a stimulus is presented, the harder it is for it to develop associations to other stimuli as either an unconditioned stimulus (US desensitization effect) or a conditioned stimulus (pre-exposure effect). Interestingly, this occurs if the stimulus is presented in isolation but not if presentations of the stimulus are interspersed with presentations of a related

stimulus. Hall (2003) presented rats with neutral edible stimuli. All stimuli were mixtures of two substances. One of the substances was shared between all of the stimuli while the other component of each compound stimulus distinguished it from the others. Two stimuli (AX and BX) were presented in alternation while the third stimulus (CX) was presented in a separate block of trials. All stimuli were presented the same number of times, that is, they had the same token frequency for the subjects. The same stimulus served as B or C for different groups of subjects. All rats were then exposed to an unconditioned stimulus Y, a nauseating substance. The rats were then presented with BY and CY. Regardless of which stimuli served as B and C, the rats consumed more CY than BY. Thus the rats habituated to the stimulus that was presented in isolation (C) more than to the stimulus presented in alternation with another (B). Hall (2003) explained the finding by saying that AX activated BX when it was presented because of the similarity between AX and BX while CX was never associatively activated and that associative activation counteracts habituation while direct, perceptual activation drives it. The reason CX was never associatively activated is that the subjects either did not yet know of the existence of CX when AX was presented or CX lost associability by the time first exposure to AX occurred. These results strongly suggest that distinct type and token representations are necessary.

## ***2.2. The Dual Mechanism Model***

### **2.2.1. Basic Features**

The Dual Mechanism Model (Pinker and Prince 1988, 1994, Pinker 1999, Clahsen et al. 1992, Clahsen 1999, Berent et al. 1999) claims that there is a qualitative difference

between regular and irregular inflection in that the regular pattern is applied by a rule, understood as an operation over variables while new irregulars are formed by analogy to existing irregulars stored in an associative network. The exact nature of the associative network has never been fully specified, however it is likely that, given the inability of distributed representations to account for dissociations between type and token frequency effects, and between the effects of direct and associative activation as well as for sequential structure, a localist network is presupposed. Otherwise, the Dual Mechanism Model would be subject to many of the same criticisms that its proponents directed against connectionist models. A particularly strong version of the model has been proposed by Berent et al. (1999). While the standard Dual Mechanism Model claims that in some languages there are some inflectional domains that utilize a default operation over variables (Pinker and Prince 1988, 1994, Pinker 1999, Clahsen 1999), Berent et al. (1999) claim that every inflectional domain in every language contains a default rule. The Berent et al. (1999) model has the advantage of being more testable and avoiding the need to specify why speakers would consider a morpheme to be the default in one domain but not in another, which has never been made clear in the standard model.

### **2.2.2. Successes**

Other theories of morphological productivity explain high productivity by distribution of the stems obeying competitor patterns throughout the lexicon and the frequency characteristics of the patterns and words that contain them. The Dual Mechanism Model, on the other hand, states that default status is independent of these factors. Thus dissociations between high productivity, high type frequency, low token frequency, low

relative frequency, even distribution through the lexicon, and concatenativity would potentially offer support for the model. In actuality, however, token frequency and relative frequency have not been taken into account when cases of dissociations have been presented. Only to the extent that these confounds are absent, the observed dissociations indicate the model's success.

Four cases of dissociations between type frequency and productivity, i.e. minority defaults, have been proposed. Clahsen (1999) summarizes the evidence for minority default status of German –s plurals and past participles. McCarthy and Prince (1990) describe a similar situation for Arabic plurals. Berent et al. (1999, 2002) present a case for a minority default in the Hebrew plural system.

### **2.2.3. Criticisms**

Bybee (1995) challenged the type frequency calculations that suggested that the German -t participle is in the minority. Clahsen et al. (1992) calculated type frequencies by summing up a sample of the most frequent types belonging to each of the patterns including both derived and non-derived stems. Bybee (1995) has challenged the frequency calculations for participles by pointing out that prefixed verbs instantiating a pattern might not contribute to a pattern's productivity. Bybee and Pardo (1981) conducted a nonce-probe elicited production study in which they presented subjects with Spanish infinitives and the first person singular indicative form and asked them to produce the subjunctive (e.g. *denir*, *dengo*, subjunctive \_\_\_\_\_) in sentential contexts. They noted that three phonologically defined classes of verbs took a velar before the final vowel. One of these classes (the 'conocer' class) has the most members. The 'tener' class

has three non-derived members and many members derived from them through prefixation (e.g. *obtener*). The 'caer' class has more non-derived members than the 'tener' class but no derived members, thus its overall type frequency is lower. Nonetheless, the 'tener' class is less productive than the 'caer' class. Bybee and Pardo interpreted this result as indicating that the derived verbs do not influence the category assignment of nonce verbs and do not contribute to a pattern's type frequency (cf. Bybee 1995, 1999). However, it is not clear how this idea follows from Network Theory, which claims that all forms with sufficient token frequency are stored and that generalizations are made over stored words to derive the schemas of competing patterns (Bybee 1988, 1995, 2001), and the results might be due to the fact that all stimuli used by Bybee and Pardo were phonologically similar to non-prefixed but not to prefixed verbs. A more serious objection is that counting only the most frequent stems underestimates the number of regulars in the subjects' lexicons (Bybee 1995). Boudelaa and Gaskell (2002) have challenged the minority status of the regular Arabic plural on the same grounds.

Koepcke (1998) argues that the German *-s* plural should not be considered a default because it is not overgeneralized more than other plural affixes based on his reanalysis of the Clahsen et al. (1992) data. Similarly, McClelland and Patterson (2002) criticize the Dual Mechanism Model based on lack of a sudden increase in overgeneralization, which would indicate the acquisition point of a symbolic rule. However, Koepcke's argument is problematic because the Dual Mechanism Model suggests only that the default will be the most frequently applied competitor with stems *similar to no existing stems*. Since *-n* has a higher type frequency it is predicted that it will be overgeneralized more because many stems whose plurals are unknown to the

child will have neighbors that take  $-n$ . The Dual Mechanism Model can only be tested with stems that do not have a skewed distribution of the various patterns in their neighborhoods. McClelland and Patterson's argument is not decisive because 1) the amount of overgeneralization is small ranging from 2 to 12 percent in English (Marcus et al. 1992), hence the exact point at which the symbolic rule is formed may not be detectable; while low percentages of overgeneralization are not problematic for the Dual Mechanism Model as long as irregular past tenses for particular verbs are learned easily (Pinker and Prince 1994); and 2) Roeper (1999) proposes that children's minds contain multiple grammars at the same time with the child going through many possible grammars within a single day; therefore, the child might alternate between different defaults with one default gradually becoming considered a default more and more frequently as it is encountered in more contexts characteristic of default use.

Data that is problematic for any model that, like the DMM, relies on maximally general, non-lexically-specific rules is that children's early constructions appear to be verb specific. Thus Savage et al. (2003) find syntactic priming of passives and ditransitives only when the prime and the target share the verb with young children, while the priming is verb-independent for older speakers. A maximally general rule is independent of lexical items and thus has to be acquired simultaneously for all verbs.

Full-form token frequency effects have been observed for regular inflected forms. For instance, Stemberger and MacWhinney (1986, 1988) found that in an elicited production task in the English past tense domain no-change errors on regular verbs ending in /t/ or /d/ were more likely if the verbs were of low frequency. Alegre and Gordon (1999) found that high frequency regular past tense forms are recognized more

quickly than low frequency ones. Sereno and Jongman (1997) obtained the same finding with English nominal plurals. Hare et al. (2001) found that when English speakers are asked to write down words whose phonological shape is ambiguous between a past tense verb and some other lexical item (e.g. *aloud/allowed*) they tend to write down the more frequent homonym even if it is the regular past tense. Bybee (2000) found that rates of word-final t/d-deletion were higher for high frequency regular past tenses than for low frequency ones. While these findings do suggest that high frequency forms are stored even if they are regular, this has not been denied in the Dual Mechanism Model (e.g. Pinker and Prince 1994).

Clahsen (1999) found that the magnitudes of morphological and identity priming were the same for regular German participles and plurals but not irregular ones even though both regular and irregular inflection is done via affixation in the two inflectional domains in the language. He argued that only stems of regular forms were stored as distinct lexical entries. However, the finding was due to the fact that identity priming produced more facilitation for regulars than for irregulars, while the amount of facilitation produced by morphological priming relative to the ‘unrelated prime’ condition was identical.

The differences in identity priming are expected if irregular past participles and plurals have higher token frequencies than regular ones since high frequency words prime themselves less than low frequency words (Scarborough et al. 1977, Jacoby and Dallas 1981, Jacoby 1983, Forster and Davis 1984, Norris 1984, Jacoby and Hayman 1987, Nevers and Versace 1998, Versace 1998, Perea and Rosa 2000, Versace and Nevers 2003). Irregular inflected forms are furthermore expected to be more frequent because

high token frequency allows forms to maintain their irregularity (Bybee 1988, 1995, 2001; Pinker and Prince 1994).

The lack of differences in the amount of morphological priming suggests that high frequency of irregular inflected forms was counteracted by low frequency of their corresponding stem forms while the reverse was true for regulars because, other things being equal, words with high token frequency are primed by and prime their morphological, phonological, semantic, and orthographic neighbors more than do low frequency words. The effect holds across tasks for both inhibitory and facilitatory priming (Schuberth and Eimas 1977, Becker 1979, Schuberth et al. 1981, Stanovich and West 1981, 1983, Stanovich et al. 1981, West and Stanovich 1982, Goldinger et al. 1989, Neely 1991, Moder 1992, Borowsky and Besner 1993, Luce et al. 2000, Plaut and Booth 2000). If that is the case, then the frequency of inflected irregular forms in German must be higher than the frequency of regular inflected forms relative to their corresponding stem forms.

Since Hay (2003) shows that higher relative frequency of forms derived with an affix decreases the productivity of the affix due to the lower likelihood of segmenting the affix out of the speech stream, the observed dissociation between type frequency and productivity in German inflection does not mean that type frequency does not influence productivity. Rather, type frequency is only one of several factors and is in this case in competition with these other factors, including relative frequency, and is overridden by them. The same case can be made for the Hebrew plural (Berent et al. 1999).

Gordon and Alegre (1999) and Ullman (1999) claim that low frequency of the stems to which the regular inflection is applied is indicative of its status as a default.

However, that implies that the past tenses of low frequency stems are not stored. Stark (1997) and Stark and McClelland (2000) looked at the effects of identity priming on words, non-words composed of existing syllables and unpronounceable non-words. There was more priming for low frequency words than for high frequency words and more priming for non-words composed of existing syllables than for low frequency words. By contrast, very little priming was observed with unpronounceable non-words. Identity priming is observed to behave like similarity-based priming, exhibiting rapid decay and small magnitude, when token frequency effects are not observed, that is, when it occurs on a sublexical level, for instance, due to an extremely short prime duration (Versace and Nevers 2003). Full-fledged identity priming of pronounceable non-words implies that, unlike unpronounceable nonwords, they are stored in the lexicon and are primed on a lexical level. That would mean that the frequency threshold for storing a form is extremely low and that therefore low frequency inflected forms are likely to be stored.

An additional confounding factor present in the German cases (Clahsen 1999) and in the Arabic case (McCarthy and Prince 1990) is that the regular pattern is the only one that always preserves the stem (Bybee 1995). Dabrowska (2001) found that in Polish nominal inflection the Genitive plural features a pattern with many of the characteristics of a rule-based default but the Genitive singular does not, and it is only in the Genitive plural that all competitors except the 'regular' affix introduce changes into the stem. This dissociation is found despite similar type frequency relations between the competitors in the singular and in the plural.

While Bybee and Newman (1995) found that the speed of acquiring the novel plural patterns and final level of acquisition were the same, regardless of whether the

plural was expressed as a suffix or a vowel change, two issues must be noted: 1) English speakers might be less familiar with  $\emptyset$ -vowel distinctions signaling singular-plural distinctions than they are with vowel changes signaling singular-plural distinctions, and 2) all of the stimuli Bybee and Newman used were monosyllabic. In chapter 3, we will see that when adult native English speakers are asked to rate pairs of phonotactically legal nonsense words on how similar they sound, words that differ in phonemes adjacent to word boundaries are perceived to be less similar than words that differ in word-internal phonemes (what Aitchison, 1987, called the “bathtub effect”) if the type of difference and number of mismatched phonemes are controlled. However, this difference is not observed with monosyllabic stimuli (Kapatsinski 2004, Steriade 2004).<sup>3</sup> It might be that monosyllabic stimuli are too short for such primacy/recency effects to emerge and the confound is only observed with longer stimuli such as those used in the Dabrowska (2001) study.

The two defining characteristics of a rule-based default are 1) the inflectional domain contains one pattern that applies to nonce stems similar to no existing stems as readily as it applies to nonce stems similar to existing stems that combine with the pattern, i.e. the regular pattern does not exhibit a similarity effect, and 2) this pattern is more likely to apply to nonce stems similar to no existing stems than any other pattern (Pinker 1999, Clahsen 1999, Berent et al. 1999, 2002).

Dabrowska (2001) has demonstrated that in two inflectional domains in a single language that are comprised of the same set of stems, one domain features a default while

---

<sup>3</sup> However, it should be noted that the studies that examined the issue with monosyllabic words did not use the same task as studies that found the effect with bisyllabic verbs. Namely, the AXB task and corpus analysis of imperfect primes were used with monosyllabic stimuli and sound similarity judgments were used with bisyllabic stimuli. Thus, the differences may be due to the differences in methodology.

the other does not. This finding contradicts Berent et al's (1999) claim that all inflectional domains in all languages feature a default. Even more damaging is evidence presented in Albright and Hayes (2003) where it is shown that even the regular English past tense allomorph, -ed, displays similarity effects. If this paragon of defaultness can display similarity effects, the likelihood of finding a similarity-insensitive default affix in any language is rather low. Accordingly, McClelland and Patterson (2002) point out that the German -s plural claimed to be a minority default is the most commonly used plural affix in only a subset of the contexts in which the English -s plural is the most common plural affix. Bybee (1995) argues that it is used with stems ending in a full vowel more often than with other stems and is therefore sensitive to phonological characteristics of the stimulus. It remains to be demonstrated, however, that its sensitivity is to the presence of possible analogical models in the nonce stimulus's neighborhood, rather than final-segment-suffix associations, the latter being compatible with being always applied by a rule.

In this paper, I report on a derivational domain in Russian. For labial-final, coronal-final, and palatal-final roots, the competing affixes are all concatenative and require no changes to the root. All of the affixes are used to derive a verbal stem from a nominal root. There are two affixes that have the highest type frequency: -i- and -a-. Interestingly, -i- is the most frequent suffix after coronal- and palatal-final stems while -a- is the only affix that does not exhibit similarity effects. Thus the defining characteristics of a rule-based default are dissociable (see chapter 4 for details and chapters 5-7 for the account of what influences the size of similarity effects and why productivity depends on the place of articulation of the root-final consonant).

## ***2.3. Rule-Based Learner***

### **2.3.1. Basic Features.**

The Rule-Based Learner was proposed by Albright and Hayes (2002, 2003) to account for the presence of similarity effects for both regular and irregular patterns even in the English past tense. The model proposes that humans internalize a system of a large number of redundant rules as a result of an inductive learning process. In this framework, multiple rules correspond to the same operation, e.g. ‘add –ed’, but all rules have different phonological conditions of application, e.g. the rules ‘add [t] after [k] to form the past tense’, ‘add [t] after voiceless velar consonants to form the past tense’, ‘add [t] after voiceless consonants to form the past tense’ may exist at the same time. Rules are kept in the grammar depending on their reliability, calculated as the proportion of word types in the lexicon that fall under the rule’s conditions of application that actually obey the rule with a built-in bias that favors rules that can apply to many word types.

Albright and Hayes’s model predicts similarity effects because the greater the reliability of a rule in a given neighborhood, the more likely it is that it will remain in the grammar after a more general rule has been induced and the more likely it is to apply in the future to verbs in the neighborhood. The difference between Albright and Hayes’s model and an analogical model is that conditions of application of Albright and Hayes’ rules are necessary and sufficient for the rule to apply in that all members of the set to which the rule applies must be similar in exactly the same way. On the other hand, forms in analogically formed gang may be similar to each other in different ways with some forms sharing final phonemes, some sharing initial ones, and some sharing medial ones

(what Albright and Hayes call variegated similarity).

A precise algorithm for calculating phonological similarity between any two words is proposed. When several affixes compete for attachment to the stem, the speaker compares the phoneme located inside the nonce stem in the position adjacent to the affix boundary with the phonemes located in the same positions in existing stems. If the phonemes match, the speaker compares phonemes in the second position away from the morpheme boundary and so on. If there is a mismatch, the number of features shared between the mismatched phonemes is calculated but phonemes that are further away from the morpheme boundary than the first mismatched phoneme are not taken into account in calculating the similarity between the words. If the affix in question is an infix or part of a discontinuous pattern fitting into a discontinuous root like in Arabic or Hebrew or a stem-internal phoneme change, the phoneme comparison process proceeds outward from both sides of the affix/pattern in parallel. This structured progression ensures that phonemes that are close to the morpheme boundary will be assigned more weight in determining the likelihood of application of a given rule to a given stem than phonemes further away from the morpheme boundary.

### **2.3.2. Successes**

The model is able to account for similarity and type frequency effects. Albright and Hayes (2003) tested the model against an analogical model. In order to test the models, the authors used an implementation of their rule-based model to create 1) a set of nonce phonotactically legal English verbs similar to existing regular verbs (i.e. verbs that were in an island of regularity for the regular rule), 2) a set of verbs that were similar to

existing irregular verbs (i.e. verbs that were in an island of regularity for an irregular rule), 3) a set of verbs that were similar to both many regulars and many irregulars, and 4) a set of verbs that were similar to neither regulars nor irregulars. In addition, a set of verbs that were extremely similar to a small number (often one) very frequent irregular verb was constructed to test for the existence of single-form and few-form analogies.

In the first experiment, native English speakers were asked to produce past tenses of the verbs given the stem form. The stem forms were presented auditorily in a sentential context and subjects filled in a blank in a third auditorily presented sentence. In addition, they rated stem forms' phonological well-formedness on a seven-point scale ranging from 'completely bizarre, impossible as an English word' to 'completely normal, would make a fine English word'. All verbs used in the experiment were rated significantly higher than phonotactically ill-formed verbs. In the second experiment, they were asked to produce the past tense form and then to rate the goodness of a regular past tense and one or two irregular past tenses as past tenses for the verb. A seven-point scale ranging from 'completely bizarre, impossible as the past tense of the verb' to 'completely normal, would make a fine past tense for the verb' was used.

Production probabilities and ratings were highly correlated. Similarity effects were found for both regulars and irregulars in both ratings and production probabilities contrary to the claims of the Dual Mechanism Model. The existence of similarity effects suggest that the regular rule's condition of application are not completely open as believed by DMM's proponents.

Computer implementations of Albright and Hayes' rule-based model and an analogical model were used to generate ratings for every regular and irregular form for

every word used in the experiment. Both models were exposed to the same training set consisting of several tens of thousands of English verbs. Frequency of exposure to each verb was proportional to its frequency. The models' ratings were then compared to the ratings given by the subjects. The correlation between the rule-based model's ratings and participants' ratings was higher than the correlation between ratings of the analogical model and ratings of the participants. Examination of results for individual items revealed that this was because the analogical model underpredicted acceptability of forms in islands of reliability. The analogical model depended on single-form analogies more than the rule-based model and the participants. Finally, the analogical model frequently failed to correctly predict which allomorph of the regular affix would be used, an error that was very rare in the human data. Unlike the analogical model, the rule-based model was able to locate the crucial dependence between the allomorph and the final consonant of the stem.

### **2.3.3. Criticisms**

Albright and Hayes argue that the problem with the analogical model is its reliance on variegated similarity. However, another difference between the rule-based model and the analogical model is that mismatches in any position receive the same weight in the analogical model. By contrast, the rule-based model is biased to focus on phonemes that are closer to the boundary between the stem and the past tense affix. These are the phonemes that are most relevant for affix and allomorph selection. This bias is not an automatic consequence of structured similarity. Neither is its absence a necessary feature of an analogical model. For instance, connectionist networks automatically increase

weights of features that are more informative for the task at hand.

In addition, the rule-based model rewards contiguous matches while the analogical model does not. Kidd and Watson (1992) find that several short mismatches between a pair of tone sequences are more salient than a single long mismatch of the same duration as the short mismatches combined. These results indicate that continuous mismatches are less salient even in non-linguistic perception. There is nothing inherent to an analogical approach that would be incompatible with this feature of the perceptual system. In fact, since non-continuous mismatches are likely to disrupt more perceptual chunks it would in fact predict this difference as long as chunking is allowed.

Another problem with the analogical model is that token frequency is assumed to make a word a better model. Hence, the overreliance on single-form analogies based on high-frequency words observed when testing the analogical model against subjects' ratings. Since high token frequency reduces the likelihood that a word will serve as an analogical model for humans (Moder 1992) and Albright and Hayes' analogical model assumes the opposite, overreliance on single-form analogy is expected but may not be due to the model's reliance on analogy.

Both models were tested on how well they were able to predict the best islands of reliability found by the rule-based model. They were not tested on islands discovered by the analogical model. It might be the case that the rule-based model would not do as well at discovering the best islands discovered by the analogical model but we have no way of knowing from the results. However, even when the results presented are submitted to a linear regression, the analogical model accounts for extra variance over the rule-based model. Thus, they must both be capturing some of the relevant characteristics of the

stimuli.

Furthermore, there is much variation between participant ratings of words in the same experimental condition (e.g. the subjects' ratings of regular forms divided by ratings from irregular forms for verbs that are claimed to be in an island of reliability for irregulars but not for regulars ranges from  $4.22/6.09=0.69$  to  $4.17/2.74=1.52$ ). Since the islands used to derive the materials were discovered by the rule-based model, their inconsistency reflects on the rule-based model and leaves open the possibility that an analogical model would be able to discover more consistent islands.

Even with all of the imperfections of the analogical model, its ratios of regular to irregular ratings are closer to the subjects' ratios than are those derived by the rule-based model for 46% of the 75 stimuli. Thus the findings do not unambiguously support the conclusion that the analogical model failed to perform as well as the rule-based model did because it relied on variegated similarity.

While the similarity-calculating algorithm used by the Rule-based Learner allows the model to zero in on the phonemes in the stem that are most relevant for competitor morpheme and allomorph selection, it also predicts that the only two factors that influence how similar two words are perceived to be are 1) how close the nearest mismatched phonemes are to the morpheme boundary, and 2) how similar the mismatched phonemes nearest to the morpheme boundary are. That means that the number of mismatched and matched segments in the two words has no effect on how similar they are perceived to be.

However, Vitz and Winkler (1973), Derwing and Nearey (1986, 1995), and Bendrien (1992) found that when English speakers are asked to rate nonsense words on

how similar they sound, words that differ by one phoneme in the word-final position are rated as more similar than words that differ by two phonemes one of which is in the word-final position even if words differing by one phoneme differ by more features. Kidd and Watson (1992) have examined native English speaker discrimination of tone sequences and found that the salience of the difference between two-tone sequences depended mostly on what proportion of the total duration of each of the tone sequences was mismatched. In chapter 3, we will report a replication of the finding with bisyllabic nonsense words (see chapter 3 for details).

While these results are inconsistent with Albright and Hayes' algorithm, it might well be that the way phonological similarity is estimated differs across tasks. In particular, the Kapatsinski (2004) results shown in (3) can be explained by a more general version of the Albright and Hayes algorithm where two words are compared in parallel starting from both ends. It is plausible to assume that the comparison process proceeds from both ends towards the middle in pure sound similarity judgment tasks and that one end is chosen whenever words are compared for the purposes of inflectional or derivational class assignment.

(3) Mismatch salience: Word final >> word-initial >> word-medial C in a word-initial  
On > word-medial C in a word-medial On

Thus to disprove Albright and Hayes' algorithm, one must use morphological productivity in an elicited production or rating task as a dependent variable. An opportunity for conducting such a test is presented by Spanish irregular verbs examined by Bybee and Pardo (1981). The 'caer' class has more non-derived members than the 'tener' class but no derived members, thus its overall type frequency is lower.

Nonetheless, the 'tener' class is less productive than the 'caer' class.

However, derived verbs were one syllable longer than Bybee and Pardo's stimuli and hence might not be sufficiently phonologically similar to them to influence the stimuli's category assignment. If this is true, derived verbs would influence the category assignment of nonce stimuli that are similar to them and should be included in type frequency counts. This hypothesis assumes that the first syllables of stems influence their phonological similarity even if there is a mismatch in a phoneme that is closer to the end of the word. That is, the hypothesis can only be true if morphological categorization is accomplished on the basis of variegated similarity (Albright and Hayes 2003).

If the subjects use Albright and Hayes' similarity algorithm, there should be no differences in the productivity of the three velar-inserting classes for stimuli similar to derived and non-derived verbs. If they take all phonemes in the word into account, the 'tener' class should be less productive than the 'caer' class for verbs similar to non-derived verbs, while the opposite should be true for stimuli similar to derived verbs. Note that if this is true, the criticism of type frequency calculations of the German participles proposed by Bybee (1995, 1999) does not apply.

The test used in this paper is to see whether root bodies can form associations with suffixes in Russian. Albright and Hayes' metric predicts that this should not be possible in any language. We find that such connections do emerge. The results are reported in chapter 7.

Using variegated similarity does not mean weighing all features in all positions equally. At least three reasons exist for why speakers might zero in on stem-internal phonemes adjacent to stem boundaries as likely determinants of morpheme and

allomorph selection: 1) phonemes adjacent to unit boundaries are more salient because they are less subject to co-articulation effects and because of the recency/primacy effects (what Aitchison 1987 termed the “bathtub effect”), 2) adjacent phonemes, including phonemes spanning a morpheme boundary, are easier to chunk into a unit because they are more likely to be activated simultaneously and to be co-activated longer; if connections strengthen based on frequency and duration of simultaneous activation, relations between adjacent phonemes will be identified more easily, and 3) various features and feature combinations differ in informational load depending on position within a word in that the same featural difference in different positions may distinguish between different numbers of words leading to position-dependent weighting. The differences between weighted variegated and structured similarity are 1) whether all phonemes within the compared words are taken into account when similarity between them is calculated, and 2) whether classical categories are stronger than radial ones<sup>4</sup>, not whether all phonemes are weighed equally.

An additional problem, noted by Albright and Hayes (2003) is that the model is entirely source-oriented, thus product-oriented generalizations are not captured. Koepcke (1998) reanalyzed data from a number of nonce-probe elicited production studies of nominal plural formation in English and German as well as naturalistic observation data and found that the likelihood of using the singular form as a plural depends on how much the singular resembles prototypical plurals and how little it resembles prototypical singulars. Wang and Derwing (1994) found that the probability of using a vowel change to form the past tense in English depends not on the frequency of the rule but on how

---

<sup>4</sup> Classical categories are categories that have necessary and sufficient features, while a radial category is a category in which every member of the category shares something with at least one other member of the category but there need not be any attributes shared by all members of the category (Lakoff 1987).

frequent the vowel produced by the rule is in the past tense forms stored in the lexicon. Albright and Hayes (2003) propose to handle these cases by surface constraints on phonology but it is not clear how this can be accomplished since the variation is morphologically, rather than phonologically conditioned. Thus, the same vowel might be a good marker of the past tense but a poor marker of the plural.

## ***2.4. Network Theory***

### **2.4.1. Basic features.**

Network theory (Bybee 1985, 1988, 1995, 2001) proposes that inflected forms are formed by analogy to forms that already exist in the lexicon which is viewed as a localist associative network. To say that it is localist means that different words are represented by separate nodes or groups of nodes and not as overlapping patterns of activation over a single group of nodes. Similarity is represented as strength of links connecting the words rather than as overlap. Token frequency is reflected in the resting activation levels of various words. Words are proposed to have phonetically explicit representations but not every token is proposed to create a new representation. Rather, tokens are mapped onto more abstract exemplars (Bybee 2001). Categories are proposed to be clusters of exemplars. Whole words are proposed to be stored in the lexicon, while morphemes and other sublexical units are generally not considered to have separate representations. High token frequency is proposed to weaken the connections between the frequent unit and its neighbors (Bybee 1985: 118,123-4, 2001, Moder 1992). Strength of a particular competitor pattern is proposed to be a function of 1) the morpheme's type frequency (Bybee 1995), 2) the morpheme's token frequency/type frequency ratio (Moder 1992,

Bybee 1995), 3) average token frequency of the complex form relative to the token frequency of the base for all words containing the morpheme (relative frequency, Hay 2001, 2003), 4) segmentability of the morpheme, as measured by transitional probability across the morpheme/stem boundary (Hay 2003) and perceptual salience of the morpheme (Koepcke 1998, Dabrowska 2001), and 5) number of competitors of the same type (Bybee and Newman 1995).

The degree to which each of the factors proposed within usage-based theory influences productivity can be determined through the use of an artificial grammar learning task of the type used by Bybee and Newman (1995). However, an important issue that has to be confronted is whether the same areas of the brain are activated when an artificial grammar is acquired in adults as when children acquire the grammar of their first language. If not, the task might have limited relevance to the question of how children determine the relative productivity of synonymous morphemes, and it might be more appropriate to test young children using less demanding testing conditions. While the influences of the proposed factors have yet to be decoupled, it is clear that at least some of them play a role in determining productivity, that at least some generalizations are product-oriented, and that semantic and phonological similarity influence morpheme choice.

#### **2.4.2. Successes.**

The approach is able to predict type and token frequency effects and does not confound type and token frequency, correctly predicting that type frequency increases productivity while token frequency decreases it (Moder 1992). It also correctly predicts that the

likelihood of using a pattern with a nonce word depends on what real words the nonce stimulus is similar to. By using localist representation Network Theory can represent sequences with repetition.

Hay (2001, 2003) proposed that it is not the token/type ratio but the average frequency of the complex forms instantiating a pattern relative to their bases that influences the pattern's productivity. She compared several English derivational suffixes and prefixes (*trans-*, *ex-*, *-ness*, *mis-*, *-less*, *in-*, *dis-*, *-ly*, *-ful*, *un-*, and *-ment*) and found that the lower the average relative frequency of a pattern was, the higher was its productivity. However, the affixes were non-synonymous and productivity was not measured with nonce stimuli similar to no existing words but by the proportion of words containing the affix that occurred only once in a corpus to total number of words containing the affix. This method does not control for whether the stimulus is stored in the lexicon together with the affix as well as what the distribution of words bearing the competing affixes is in the stimulus's neighborhood. It remains to be seen whether the correlation would hold for synonymous affixes and novel words similar to no existing words. As noted before, relative frequency may be able to explain the high productivity of the German -s plural relative to what its type frequency would predict (Hay 2001). However, this hypothesis has not yet been tested.

Hay (2003) proposed that the more segmentable a morpheme is from the words that contain it, the more productive it is. Segmentability includes not only relative frequency but also phonotactic probabilities at the morpheme's boundaries. In the same study of derivational morphemes in English, Hay demonstrated that more productive morphemes tend to have low-probability transitions at the morpheme-stem boundary and

thus are easily segmentable from the stem. Koepcke (1998) proposed that the salience of morphemes is also important and low salience of umlaut can contribute to its relatively low productivity as a plural marker in German. However, as Koepcke himself points out, type frequency would also predict low productivity of umlaut in that domain.

Strong evidence for number of similar competitors as a determinant of productivity comes from Bybee and Newman (1995) who conducted an artificial grammar learning experiment in which adult native English speakers learned two to four novel plurals, some of which were realized as a vowel change in the stem, while others were realized as a single-vowel suffix. Subjects were divided into 4 groups. Each group was exposed to singular-plural pairs that exhibited both plural suffixes and plural stem changes. For all groups, the lexical distribution of the morphemes was phonologically and semantically unpredictable. The type frequency/token frequency ratio was controlled by presenting each singular-plural pair the same number of times. For one group, there was one suffix and one stem change, for another there were four suffixes and one stem change, for the third group there were four stem changes and one suffix, and for the fourth group there were two suffixes and two stem changes. All roots were monosyllabic nonsense words. After the subjects were exposed to the training set and practiced producing the plural given the singular with feedback, they were exposed to unfamiliar singulars and asked to produce the plural. In this generalization stage, whichever location-specific morpheme type, whether stem change or suffix, contained only a single competitor was more productive.

Summed type frequency of stem changes and suffixes was always the same and did not differ across conditions in Bybee and Newman (1995). However, if the effect of

type frequency on productivity is nonlinear, type frequency may still be able to account for the results. Bybee and Pardo (1981) proposed that morphemes with a type frequency of three or below are not productive in natural languages. More generally, Bailey and Hahn (2001) hypothesize that the same increase in type frequency is more salient at medium frequencies than at high or low ones. If that is the case, and the morphemes in the 4-competitor condition are of sufficiently low type frequency, Bybee and Newman's results may be predicted by type frequency.

Bybee and Moder (1983) and Koepcke (1998) have proposed that cue validity of the morpheme is important in that a polysemous or homonymous phonological shape might be harder to acquire and thus less productive than a monosemous one. Koepcke (1998) proposes that the plural morphemes *-e* and *-er* in German have low cue validity, which makes them less productive than would be expected if productivity was determined by type frequency alone. In addition, he proposes that *-s* has lower cue validity as a plural marker in English than *-z* or *-əz* and that for this reason children tend to interpret singular forms ending in /z/ or /əz/ as plural more frequently than singulars ending in /s/. Finally, he shows that the distance of the word-final consonant in the singular from the plural markers is predictive of how likely the singular is to be misinterpreted as a plural. This finding is argued to support the notion of cue validity because subsegmental feature clusters the root-final consonants contain could also be more or less valid cues of plurality. While the evidence is suggestive, the hypothesis has to be subjected to more systematic testing.

Network Theory is able to predict that less priming should be observed with high-frequency primes and targets in similarity-based and co-occurrence-based priming (see

Moder 1992 for prime frequency in morphological priming, Thomsen et al. 1996 for prime frequency in semantic priming, Schuberth and Eimas 1977, Schuberth et al. 1981, Becker 1979, Stanovich and West 1981, 1983, Stanovich et al. 1981, West and Stanovich 1982, Neely 1991, Borowsky and Besner 1993, and Plaut and Booth 2000 for target frequency in semantic priming, Perea and Rosa 2000 for target frequency in orthographic priming, Goldinger et al. 1989 and Luce et al. 2000 for prime frequency in phonological priming). However, it so far has not said anything about the reasons for the absence of the prime frequency effect in orthographic priming (Segui and Grainger 1990, Drews and Zwitserlood 1995, Bijeljac et al. 1997, Brysbaert et al. 2000).

Network Theory is also supported by the existence of product-oriented generalizations. Lobben (1991), summarized in Bybee (1995), argues for a product-oriented analysis of Hausa plurals based on low predictability of the plural from phonological and semantic characteristics of the singular and the fact that the characteristics of the plurals can be captured by fewer product-oriented generalizations than source-oriented ones. Wang and Derwing (1994) provide strong evidence for product-oriented generalizations by showing that the productivity of irregular past tense morphemes in English depends on the frequency of the stem vowels as past tense markers and not on the frequency of the putative operation converting the present tense vowel into the past tense vowel. In addition, Wang and Derwing (1994) found that when native English speakers are asked to form nouns with *-ity* from nonce adjectives, the vowels used in the stem of the produced noun are predicted by their type frequencies in nouns ending in *-ity* but not by the vowels of the adjectives from which the nouns have been formed. Further evidence of product-oriented generalizations is provided by no-change

errors in English past tense and plural formation, which are more likely to occur the more features of the final consonant of the stem and the penultimate vowel typically occur in word-final markers of plurality and past tense. Thus, Koepcke (1998) shows that singulars ending in /ɪz/ are more likely to be left uninflected in the plural by children learning English than singulars ending in /s/, which are more likely to be left unchanged than those ending in /ʒ/ or /tʃ/. Even less likely to be uninflected are stems ending in /f/, which nonetheless are unchanged more often than stems ending in a vowel, /r/, /n/, or /g/.

### **2.4.3. Criticisms**

High type frequency has been proposed by Bybee (1995) to lead to high productivity. However, the degree to which type frequency influences productivity is still unclear. In both the Guillaume (1927/1973) and the Bybee and Pardo (1981) studies high type frequency is confounded with a low token frequency/type frequency ratio. In Wang and Derwing's study similarity of nonce stimuli to existing words in the lexicon was not controlled. In Bybee and Pardo's study no information is given about how the degree of similarity of the nonce stimuli to the potential analogical models was controlled. Finally, the studies reviewed above provide only correlational evidence and reverse correlations between productivity and type frequency have also been reported for German (Clahsen 1999), Arabic (McCarthy and Prince 1990) and Hebrew (Berent et al. 1999) plurals. Baayen and Lieber (1991), while finding a correlation between type frequency and productivity of *-ness* vs. *-ity*, observe that the difference in type frequency appears too small to account for the large differences in the perceived productivity of the two suffixes.

Low token frequency to type frequency ratios have been proposed to lead to high productivity by Bybee (1985, 1995). The theoretical underpinning of this effect has been proposed to be that words with high token frequency are likely to be accessed as wholes and thus have weak connections to morphemes they contain and other words in the same gang. Thus, a given number of high frequency words may not contribute to the type frequency of a morpheme they contain as much as the same number of low frequency words. Evidence for inaccessibility of morphemes from high-frequency words that contain them relative to lower-frequency words is provided by Moder (1992). She presented native English speakers with past tense forms of high or medium frequency instantiating a particular irregular pattern in English. For a given pattern, the subjects were more likely to apply it to nonce stimuli in a subsequent elicited production task if they were primed with a medium-frequency form instantiating the pattern than if a high-frequency form had been presented. However, as was pointed out earlier, type frequency and token/type ratio have been confounded in studies examining their effect on productivity.

There is some disagreement regarding whether concatenativity plays a role in productivity. Dabrowska (2001) proposes that the reason overregularization is found for the same children in the Polish genitive plural but not in the genitive singular is that in the singular all competitors are suffixes while in the plural only the overgeneralized one is. By contrast, Bybee and Newman (1995) found no effect of whether an unfamiliar plural acquired by native English speakers in an artificial grammar learning task was expressed as a vowel change in the stem or a single-vowel suffix on the end-state productivity of the morpheme and its speed of acquisition. Bybee and Newman's (1995)

data also contradict the proposal that additive morphemes are more productive in the English plural because adding a sound or sounds is evocative of plurality (what Koepcke 1998 calls ‘the principle of iconicity’).

As it now stands, Network Theory fails to predict that as frequency increases, an increase in frequency of a given magnitude becomes more and more insignificant. That is, behavior is proposed to be sensitive to raw frequency, not logarithmically scaled frequency. This proposal is contradicted by findings of reduced identity priming with high-frequency words (Scarborough et al. 1977, Jacoby and Dallas 1981, Jacoby 1983, Forster and Davis 1984, Norris 1984, Jacoby and Hayman 1987, Stark 1997, Stark and McClelland 2000, Nevers and Versace 1998, Versace 1998, Perea and Rosa 2000, Versace and Nevers 2003).

Network Theory has no mechanism for predicting the age-of-acquisition effect, i.e. the finding that when token frequency is controlled, words acquired early in life are more accessible (Bonin et al. 2001, Meschyan and Hernandez 2002, Morrison et al. 2002, 2003, Newman and German 2002, Zevin and Seidenberg 2002, 2004, Ghyselinck et al. 2004) and are primed less in repetition priming (Barry et al. 2001).

Network Theory predicts that a word A is similar to word B to the same extent as word B is similar to word A failing to account for the differences in priming between the same pair of words depending on which is used as the prime and which as the target. In particular, 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<sup>5</sup> (Goldinger et al. 1989), and phonological (Radeau et al. 1995) priming.

Network Theory has no way of accounting for morphological priming as distinct from phonological and semantic priming. Thus, the result that phonologically and semantically related pairs of words do not show long-term priming (Stockall 2004) but morphologically related pairs do is not predicted. Furthermore, the finding that morphological priming is long-term and facilitatory while phonological priming is short-term and inhibitory and semantic priming is short-term and excitatory (Stockall 2004) is problematic for Network Theory just as it is for connectionism.

Finally, just like connectionist models, Network Theory makes no predictions regarding the effects of direct and associative activation (Hall 2003). In Network Theory, while only direct activation leads to connection weakening, associative activation does not lead to connections strengthening. Thus Hall's finding that, when number of exposures is controlled, the stimulus whose presentations are interspersed with presentations of a related stimulus does not lose associability is not predicted.<sup>6</sup>

### ***2.5. Local Activation Spread Theory<sup>7</sup>***

Since none of the existing models can explain all of the results observed, I would like to propose what we shall call Local Activation Spread Theory, or LAST. LAST is a localist implementation of Network Theory (Bybee 1988, 1995, 2001) that uses Spreading

---

<sup>5</sup> Acoustic/phonetic priming is priming between words that are confusable in noise but don't share any phonemes, e.g. *veer-bull*.

<sup>6</sup> Although, as Joan Bybee (p.c.) has pointed out, the finding needs to be replicated with auditory stimuli in humans.

<sup>7</sup> This section has been presented at HDLS VI and accepted for publication in the proceedings of the conference. It is also scheduled for presentation at LACUS XXXII: Networks.

Activation Theory (Collins and Loftus 1975, Anderson 1976, 1983, 2000) and Exemplar Theory (Hintzman 1986, Nosofsky 1988, Goldinger 1992, Johnson 1997, Pierrehumbert 2001) to provide a unified account of frequency, neighborhood density, and lexicon size effects in priming, word recognition, nonce probe tasks, and associative learning.

### **2.5.1. Basic features**

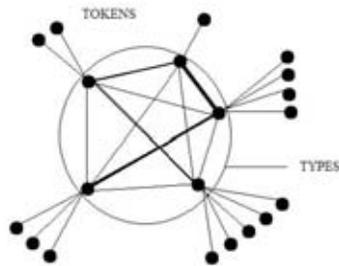
#### **2.5.1.1. Architecture**

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.

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. 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 preading 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 of the prime's presentation 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.

The architecture is presented in Figure 2.2.

**Figure 2.2. Architecture of memory according to LAST<sup>8</sup>**



### **2.5.1.2. Microstructure**

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).

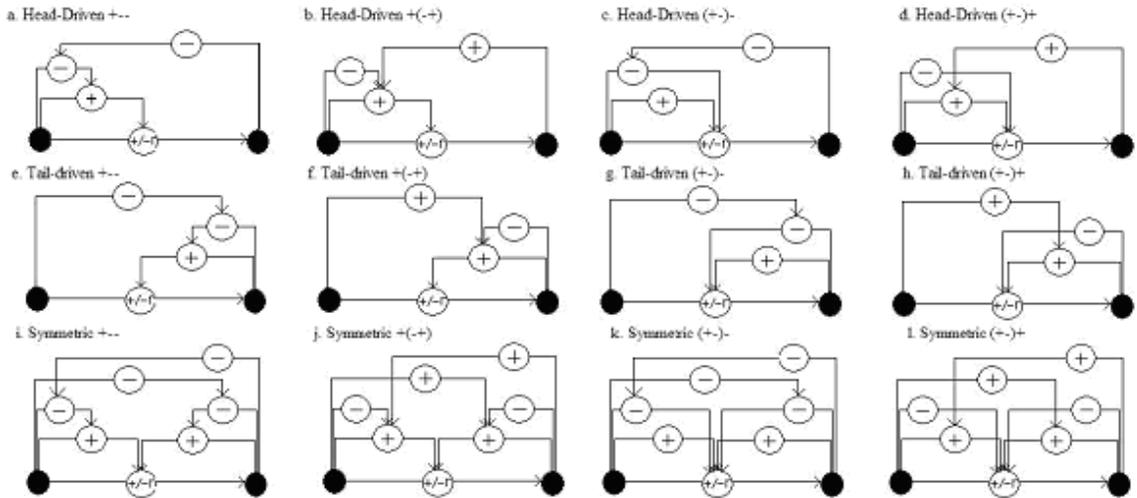
On the other hand, a link should strengthen whenever its head and tail are activated simultaneously (co-activated) to allow associative learning to occur. This would imply that the PF's r-value is raised whenever the head and the tail of the link are co-activated and hence that the PF of a link is a tail on a subsidiary link (LINKTRON) headed

---

<sup>8</sup> Lines represent connections, width of line represents connection strength, filled circles represent nodes. The circle surrounding the type nodes has no theoretical significance.

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 twelve structures are possible a priori (Figure 2.3). However, we will see in Section 2.5.1.4 that there are strong empirical constraints on possible structures.

**Figure 2.3. A priori possible LAST link structures<sup>9</sup>**



### 2.5.1.3. Dynamics of activation spread

LAST proposes that each node has an ACTIVATION THRESHOLD. After the activation 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 (Anderson 2000). LAST differs from previous semantic networks models 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

<sup>9</sup> 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.

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 activation unit is an amount of activation that has entered a particular node or link at a particular point in time. Thus 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.

#### **2.5.1.4. Empirical Constraints on Link Structure**

It has been observed that for a given pair of stimuli, different amounts of priming can be observed depending on which stimulus serves as the prime and which serves as the target (e.g. Koriat 1981, Goldinger et al. 1989, Schriefers et al. 1992, Radeau et al. 1995, Chwilla et al. 1998, Feldman 2003, Rueckl 2003). In LAST, these differences can only be due to differences in the strengths of the links connecting the stimuli in different directions, because both the prime and the target's neighborhood density and token frequency influence the amount of priming, and they influence it to exactly the same degree. If the structure of the link is symmetric, the strength of both links connecting two nodes is identical, and thus there is no explanation for asymmetric priming effects. If the structure is asymmetric, the DRIVING NODE of a link is the only one sending activation to the PF. An increase in the token frequency of the driving node would mean less activation being 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 be weaker than the link driven by the low frequency node. Since, 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<sup>10</sup> (Goldinger et al. 1989) and phonological (Radeau et al. 1995) priming, and since in priming activation must flow from the prime to the target, with prime therefore acting as the head of the link, we may conclude that links are head-driven. Thus, only a, b, c, and d (Fig.2) are possible structures.

Linktrons PF's, unlike link PF's, are non-trainable. If this were not the case, the linktron headed by the tail of the link would strengthen or weaken one of the linktrons headed by the head of the link, upsetting the equilibrium and allowing the head 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 a, b, and c are possible structures.

Linktron PF's are also binary. That is, a linktron can be either able or unable to transmit activation. If that were not the case, link strength would depend equally on the frequency of the head and the tail and priming asymmetries would be impossible. If linktron PF's are binary, excitatory linktrons tailed by linktron PF's have no effect unless they are counteracting inhibitory linktrons with the same tail. Furthermore, structures where two linktrons have the same tail but different heads (e.g. b) are highly unstable

---

<sup>10</sup> The finding thus holds across excitatory and inhibitory priming.

with the effect of the structure on the link's PF depending on, in part, the frequencies of the heads of the linktrons. Therefore, only a and c would produce priming asymmetries. 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. Let us define  $t$  as the speed of transmission of activation through a link(tron). Then we can distinguish between  $t_0$ ,  $t_1$ , and  $t_2$  link(tron)s such that  $t_n > 2t_{n-1}$ .<sup>11</sup>

### 2.5.2. The LAST account of the findings

In priming, when the prime is presented, matching types are partially activated. Whenever a type passes the activation 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 et al. 1977, Becker 1979, Glanzer and Ehrenreich 1979, McClelland and Rumelhart 1981, Schvaneveldt and McDonald 1981, Paap et al. 1987, Gordon 1983, Norris 1984, Goldinger et al. 1989, Monsell 1991, Luce et al. 2000, and Plaut and Booth 2000. Slow recognition feeds back to low experienced token frequency since slow

---

<sup>11</sup> If we want to minimize the number of speeds to be distinguished, +-- structures are dispreferred. This leaves structure c.

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 et al. (1977), Jacoby and Dallas (1981), Forster and Davis (1984), Norris (1984), Stark (1997), Stark and McClelland (2000), Perea and Rosa (2000), Versace and Nevers (2003) for identity priming, by Moder (1992) for morphological priming, by Thomsen et al. (1996) for semantic priming, and by Goldinger et al. (1989) and Luce et al. (2000) for phonological priming<sup>12</sup>.

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 Lorsch 1983, Schwantes 1985, Emmorey et al. 1995, Nation and Snowling 1998, Castles et al. 1999, Morford 2003), 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 Brown and Watson 1987, Bonin

---

<sup>12</sup> This finding is not predicted by Compound Cue Theory (Ratcliff and McKoon 1988).

et al. 2001, Meschyan and Hernandez 2002, Morrison et al. 2002, Newman and German 2002, Zevin and Seidenberg 2002, and Ghyselinck et al. 2004).

The finding that early-acquired words exhibit less priming (Barry et al. 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 activation threshold (Hall 2003).

Low frequency orthographic primes produce no inhibitory priming while high frequency orthographic primes significantly inhibit their targets (Segui and Grainger 1990, Drews and Zwitserlood 1995, Bijeljac et al. 1997, Brysbaert et al. 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 amount of priming<sup>13</sup> as a function of prime or target frequency is the same, regardless of modality (Figure 2.4). However, orthographic representations are acquired later than phonological representations, at a

---

<sup>13</sup> By ‘amount of priming’ we mean the magnitude of the change in the dependent variable (reaction time, brain wave latency and amplitude, error rate) due to related prime presentation compared to no prime at all.

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 from the prime will reach the target in orthography, resulting in greater incidence of the target not reaching the activation threshold as a result of prime presentation.

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) reviews 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 if the priming occurs at the most general type level, e.g. the type /p/ showing identity priming, rather than the type 'p/ occurring in /pit/' priming type 'p/ occurring in /peɪst/'. Which account fit the data better is an interesting empirical question that is important for further specifying LAST's global pattern of connectivity.

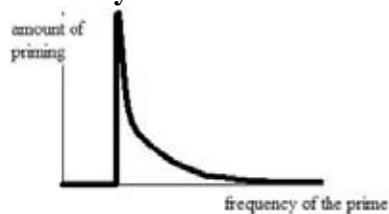
A finding that would be inconsistent with the LAST account of priming (but predicted by the activation functional account of Plaut and Booth 2000) would be if low frequency primes produced some priming but the amount of this priming were less than that produced by high frequency primes. No such evidence exists at present (both prime frequency<sup>14</sup> and target frequency, cf. Schuberth and Eimas 1977, Becker 1979, Neely

---

<sup>14</sup> See citations above.

1991, Borowski and Besner 1993, Plaut and Booth 2000, Perea and Rosa 2000, exhibit an inverse correlation with amount of priming). In addition, since LAST maintains that the modality effect in priming is due to differences in age-of-acquisition of phonological and orthographic associations and thus link strength, it predicts that late-acquired words will exhibit more identity priming (since the source node would compete with weaker links) as found by Barry et al. (2001).

**Figure 2.4. 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 activation threshold. Stark (1997) and Stark and McClelland (2000) have demonstrated that even visually presented pronounceable non-words, which are sometimes encountered across a word boundary, produce identity priming and that, furthermore, they produce even more identity priming than low frequency words.

The 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.

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. Thus, for instance, if ‘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, it is not necessary that ‘cat’ will prime ‘mouse’ as strongly as ‘hammer’ will prime ‘nail’ even if the frequencies of the prime and the target are controlled. Rather, 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, while the connection between ‘cat’ and ‘mouse’ and the connection between ‘hammer’ and ‘nail’ are of the same strength, 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 an unconditioned stimulus (US) is already strongly associated with a conditioned stimulus (CS1), it is hard to associate with another conditioned stimulus (CS2) (e.g., Kamin 1969, Marchant and Moore 1973, Kim et al. 1998). 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 pre-exposure 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, Anderson 1976).<sup>15</sup> 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.<sup>16</sup> 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. Zeelenberg (1998) failed to find the interference effect in a study in which A-C associations as well as A-B associations or only A-B associations were primed. No effect of A-C presentation on the amount of priming produced by A-B presentation was observed. However, the primed A-C associations were already strong, reducing the magnitude of the manipulation.

---

<sup>15</sup> See Whittlesea (2003) on evidence contradicting ACT's (Anderson 1976, 1983) view that the procedural/declarative distinction is a distinction between cognitive modules.

<sup>16</sup> 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.

Neighborhood effects in word recognition are inconsistent: words from dense neighborhoods are found to be recognized more slowly (e.g., Goldinger et al. 1989, Carreiras et al. 1997, Newman et al. 1997, Luce and Pisoni 1998, Vitevitch and Luce 1998, Sommers and Lewis 1999, Luce et al. 2000) or more quickly (e.g., Andrews 1989, Forster and Shen 1996, Sears et al. 1999, Huntsman and Lima 2002, Arduino and Burani 2004). 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, words that are of low frequency relative to their neighbors 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 density 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. More work is needed in this area to determine whether high frequency of the neighbors of a word would also lead to facilitatory density effects with low neighbor frequency leading to inhibitory density effects, as predicted by the theory.

In a nonce-probe task, the subject must form a new type node to represent the nonce word s/he is presented with. S/he is also asked to choose a morphological pattern

to apply to the nonce word to form a particular form of the word, e.g. its past tense. Activation then spreads outward from the new type node and also from the node corresponding to PAST TENSE. PAST TENSE would be strongly connected to all verbs in the past tense as well as to the competing morphemes. The verbs instantiating a pattern would also be strongly connected to the pattern. Therefore, the more verbs instantiating a pattern, the more activation would reach it from PAST TENSE, making its application more likely. The new type has no reason to be strongly connected to any of the past tense patterns; rather, activation flowing from it disperses across all verbs in the lexicon, with more being allocated to verbs similar to the nonce stimulus, hence, a pattern with a high type frequency will also receive more activation flowing from the nonce probe since there would be more possible mediating verbs. High token frequency of the verbs will reduce the amount of activation reaching the pattern they instantiate since the type-type links between the high-frequency verbs and the pattern would have to compete with many type-token links. Consequently, as proposed by Bybee (1995, 2001), high type frequency should increase productivity of a pattern while high token frequency of verbs the pattern occurs in should decrease it.

Unlike connectionist models, which conceive of similarity as overlap between patterns of activation (e.g., Plaut and Booth 2000), we assume that similarity strengthens connections between localist nodes, just as does co-occurrence, through co-activation.<sup>17</sup> 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

---

<sup>17</sup> Although similarity-based strengthening may occur at a different rate than co-occurrence based strengthening.

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.

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 on the other decay rapidly (cf. Stockall 2004), although long-term semantic priming has been observed by Becker et al. (1997) and Joordens and Becker (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. The source nodes in morphological priming are representations of root or even word types, which are much less frequent than the phonological units that serve as source nodes for phonological priming (cf. Radeau et al. 1995). Because of the relatively low frequency of the source node, little activation dissipates into its tokens. 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. 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.

Phonological priming of word beginnings is always inhibitory when strategic factors<sup>18</sup> are controlled (Goldinger et al. 1989, Radeau et al. 1995, Vroomen and DeGelder 1995, Luce et al. 2000). On the other hand, phonological priming of word ends is excitatory (Slowiczek et al. 1987, 2000, Corina 1992, Praamstra and Stegeman 1993,

---

<sup>18</sup> I.e., factors that are under the goal-oriented control of the subject.

Praamstra et al. 1994, Radeau et al. 1994, 1995, Cutler and Chen 1995, Radeau 1995, Titone and Connine 1997, Cutler et al. 1999, Dumay et al. 2001).

This pattern is predicted by LAST. 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, 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. This is supported by the finding that the effect of final overlap is modality-specific (cf. Radeau et al. 1995). As Radeau et al. (1995) point out, initial sounds are also more perceptually salient, and thus may have higher resting activation levels, thus priming them may not increase the speed of their recognition as much as priming final units.

One can object to the LAST account of associative learning by noting that it can only account for Hebbian learning (Hebb 1949), in which co-activation strengthens connections, and has no way to implement error-propagation. In error propagation, learning is driven by how close the activation of the output nodes is to the target level and how much each of the input nodes contributes to its deviation from the target level; if an

input node contributes much to the error, the connection between it and the target node is weakened. While error propagation has been widely used in connectionist models, no neural mechanism for it has ever been proposed (cf. Stark and McClelland 2000). In addition, there is empirical evidence against it. Baddeley and Wilson (1994), Wilson et al. (1994), and Wilson and Evans (1996) asked amnesic subjects to learn a list of morphologically complex words. Two groups of subjects were used. The ‘errorful’ group was given the stem, asked to complete the word, and later, when they guessed incorrectly three times, guessed correctly, or 15 seconds elapsed, they were shown the right word and told to write it down. The ‘errorless’ group was given the stem immediately followed by the word that was to be memorized and told to write it down. The errorless group learned to produce the words when presented with their stems better, indicating that producing incorrect guesses retards learning rather than improving it as a result of extra practice, as would be predicted if the subjects were using error-propagation. Rather, it appeared that the subjects associated incorrect responses with the stem as a result of their co-occurrence in the errorful condition, while only the correct response was associated with the probe in the errorless condition, as predicted by Hebbian learning. Thus, lack of error-propagation in LAST is likely to be psychologically realistic.<sup>19</sup>

## ***2.6. Conclusion***

Thus, LAST provides an account of frequency, neighborhood density, lexicon size, age of acquisition, target degradation, and speed of processing effects across tasks, domains, and species, using the single mechanism of spreading activation in a localist associative

---

<sup>19</sup> However, LAST can be extended in a fairly trivial way to support error-driven learning.

network, whose structure is independently motivated. In what follows, we will present further evidence incompatible with the foundational assumption of the Dual Mechanism Model and will show that the similarity-estimation method of the Rule-Based Learner needs to be modified. We will then describe some empirical ways of testing LAST in comparison to connectionist models.

## CHAPTER 3: CONTROLLING FOR SOUND SIMILARITY

There have been relatively few studies of morphological productivity that took into account the findings on what makes two words sound more or less similar to each other. However, determining that can influence a nonce stimulus's category assignment because of their phonological similarity to the stimulus is necessary to determine the degree to which the competing suffixes are sensitive to its content, i.e. the magnitude of the similarity effect associated with each of the suffixes.

### *3.1. The FIRM Method of controlling for similarity<sup>20</sup>*

The traditional method introduced by Prasada and Pinker (1993) and widely used by proponents of the DMM (e.g., Berent et al. 1999) is to take an existing word that rhymes with many members of a particular inflectional class, change it in some minimal way (e.g. by one feature) and assume that the nonce word is closest to the members of the inflectional class of the word from which it was created. The problem with this approach is that there is no quantitative control over the number of words from each of the inflectional classes that a given stimulus is close to, nor of how close it is to each of them (cf. Albright and Hayes 2003). For example, changing [baɪt] to [paɪt] makes it close to [faɪt] (past [fɔɪt]), a word of a different conjugational class than [baɪt] (past [bɪt]).

It is necessary to take into account all words that are close enough to the nonce stimulus to affect its assignment to one of the inflectional classes, i.e. all words in the

---

<sup>20</sup> This section has been presented at FASL 13 and the 4<sup>th</sup> International Conference on the Mental Lexicon (Kapatsinski 2004b) and appears in the FASL 13 proceedings (Kapatsinski 2005).

stimulus' neighborhood (Luce and Pisoni 1998) and to weigh the influence of each of these words by its proximity to the stimulus so that more similar words influence category assignment more. While it is possible that only the most similar word is taken into account as a possible analogical model (Skousen 1989), even if this is true, we would still need to know how similar the word must be to the stimulus. Furthermore, determining if only one word matters or if less similar neighbors are taken into account too requires a reliable measure of similarity and a way to derive the neighborhood given a stimulus.

The method developed for use in this study acts as a continuation of the methodological work on controlling for similarity carried out in Nosofsky (1990) and Albright and Hayes (2003) and may be termed The Fixed Radius Method, or FIRM.

Under FIRM, the experimenter first makes up a nonce stimulus, chooses a radius of X units of phonological similarity and finds all words that are within that radius of the stimulus. In this paper, a reverse dictionary of Russian (Zaliznjak 1977), and the Ogonek Corpus of modern written discourse collected in the 1990's (SFB441, 2000) were searched to obtain words in the neighborhood of a given nonce stimulus.

Second, the experimenter derives a similarity score for each of the words, which is the inverse of the number of units of similarity (UNOS) by which the word differs from the nonce stimulus. The similarity score of each of the inflectional classes in the domain is the sum of the similarity scores of its member words which are within the neighborhood radius of the stimulus. The class with the highest similarity score is the class that the stimulus is closest to. Stimuli that are not close to any existing words have no neighbors within the radius.

FIRM makes two assumptions. One is that the neighborhood is a discrete set, i.e. the set of words that are brought to bear on a word's category assignment can be delimited (cf. Luce et al. 2000). The second is that the neighborhood has a graded internal structure, i.e. that distance from the stimulus within the neighborhood, rather than simple membership or non-membership in the neighborhood is relevant. These assumptions are also made in Nosofsky (1990).

### ***3.2. Empirical findings on sound similarity***

#### **3.2.1. Previous research**

##### **3.2.1.1. Major issues and methods**

The major issues in measuring sound similarity of units that are larger than a segment are:

- 1) How does location within the stimulus impact the salience of a match (the part two stimuli share) or a mismatch (the part the stimuli do not share)? (Nelson and Nelson 1970, Derwing and Nearey 1986, Forster and Davis 1991, Bendrien 1992, Coltheart et al 1999, Kinoshita 2000, 2003)
- 2) Do characteristics of both the match and the mismatch matter? (Fallon and Robinson 1987, Kidd and Watson 1992, Schiller 2004)
- 3) What units are speaker/hearers sensitive to when assessing sound similarity?

In particular,

- a) Are they sensitive to the number of matched and/or mismatched units of some type? (McInish and Tikofsky 1969, Vitz and Winkler 1973, Jaeger 1980, Selkirk 1982, Treiman 1983, 1984, 1985, 1988, Derwing et al. 1987,

1991, 1994, Dow 1987, Johnson 1987, Treiman and Danis 1988, Derwing 1992, Beinert and Derwing 1993, Wang and Derwing 1993, Derwing and Wiebe 1994, Yoon 1994, 1995, Derwing and Wang 1995, Treiman and Kessler 1995, Bailey and Hahn 2001, Dumay et al. 2001, Yoon and Derwing 2001)

- b) Does location within the stimulus impact the salience of different units differently? (Kapatsinski 2004, Steriade 2004)
- c) Does location in relation to the unit's boundaries matter for mismatch and/or match salience? (Bendrien 1992)
- d) Are speaker/hearers sensitive to the unit's frequency and/or transitional probabilities defined at the level of this unit? (Kessler and Treiman 1997)
- e) Are the units involved abstract or acoustically specific? (Kapatsinski 2004e, 2005c, this thesis)

- 4) How does intonation structure influence the salience of mismatches and matches? (Tamariz 2005)
- 5) To what degree the salience of matches and mismatches can be explained by their acoustics, articulatory properties, and distributional characteristics. Salience is operationally defined as the impact a match or a mismatch has on the sound similarity judgments and the amount of form-based priming between a pair of stimuli (Yoon and Derwing 2001, Kapatsinski 2004e, 2005c, this thesis, Tamariz 2005).

The tasks used are:

- 1) Sound similarity judgment, or SSJ, where a pair of stimuli is presented and

subjects are asked to rate how similar the stimuli sound relative to each other on a predetermined scale (Nelson and Nelson 1970, Vitz and Winkler 1973, Derwing and Nearey 1986, Yoon and Derwing 2001, Kapatsinski 2004e, 2005c, this thesis) or in relation to the arbitrary number chosen by each particular subject to denote the perceived similarity of the first pair of stimuli (Greenberg and Jenkins 1964).

- 2) The AXB task, where subjects are presented with three stimuli and are asked whether the second stimulus sounds more like the first stimulus or more like the third stimulus (Kapatsinski 2004, Tamariz 2005),
- 3) Same-different judgments (McInish and Tikofsky 1969, Derwing and Nearey 1981, Fallon and Robinson 1987, Kidd and Watson 1992).
- 4) Form-based priming where stimuli are presented one-by-one and subjects are asked to perform some task on the stimuli; while the task is usually unrelated to sound similarity judgment, the experimenter can judge how similar the stimuli are by comparing how quickly and accurately the subjects respond to a stimulus when it is preceded by a related versus an unrelated stimulus. The greater the difference between the two conditions, the more priming there is and the greater the difference in relatedness to the target between the unrelated primes and the related primes, hence the more similar the related primes are to the target (Forster and Davis 1991, Connine et al. 1993, Radeau et al. 1995, Schiller 1998, 2004, Peresotti and Grainger 1999, Kinoshita 2000, 2003, Dumay et al. 2001).
- 5) Concept formation where subjects are asked to classify stimuli according to a criterion given by the experimenter, e.g. a shared rime (Jaeger 1980, Derwing and

Wang 1995)

- 6) Unit counting, where the subject is asked how many ‘sound units’ the stimulus consists of after being shown examples with other stimuli; ability to generalize is measured; a response range is also sometimes specified (Dow 1981, however, Derwing 1987 argues that the task is especially vulnerable to orthographic contamination),
- 7) String manipulation tasks where subjects are asked to manipulate stimuli after some practice with other stimuli by substituting or deleting parts of the string or inserting pauses; units can be detected with this task by seeing whether it is easier to insert between than inside units and to delete or substitute whole units as opposed to their parts or cross-unit strings (Selkirk 1982, Treiman 1984, Dow 1987, Derwing 1987, 1992, Yoon and Derwing 2001)
- 8) Recall of sequences sharing a particular unit (Yoon and Derwing 2001) where it is assumed that the more similar the stimuli are perceived to be the easier it will be to recall them.
- 9) Corpus studies of imperfect rhymes (Steriade 2004)
- 10) Studies of cross-linguistic commonality phonological patterns (Steriade 2001).

### **3.2.1.2. Location of mismatch in relation to word boundaries**

Several studies have investigated whether location in relation to word boundaries has an effect on mismatch salience. Nelson and Nelson (1970) asked native English speakers to judge sound similarity of real, visually presented CVC words using the SSJ task. Subjects judged words sharing the final consonant to be more similar than words sharing the initial

consonant. Derwing and Nearey (1986) found the same effect with auditorily presented CVC words in English. Bendrien (1992) replicated the effect with nonsense CVC syllables that were phonotactically legal in English. Tamariz (2005) used the AXB task with visually presented CVCV and CVCCV non-words in Spanish. Like previous studies on English, she found that initial and final matches were more salient than medial matches for both vowels and consonants.

Tamariz (2005) proposes that the salience of the word-final position might be due to the need to attend to the word-final position because that is where most morphological variability occurs. This hypothesis needs further verification in predominantly prefixing languages. However, Tamariz (2005) has demonstrated that the effect is not due simply to subjects' interpretation of the final sounds as morphemes because it holds for vowels that are not morphemes in Spanish. Alternatively, the salience of word-final sounds could be a universal recency effect, which may help in the acquisition of suffixing morphology (Slobin 1973), accounting for its predominance across languages (Cutler et al. 1985, Hall 1992, although cf. Bybee et al. 1990 who failed to find the preference for verb-initial languages). However, a learnability explanation for the predominance of suffixes is not necessary because, due to high frequency of grammatical words relative to lexical ones, the probability of a grammatical word given the previous lexical word is larger than the probability of a lexical word given the preceding grammatical word, leading to greater likelihood of the grammatical word fusing with the preceding word than with the following one (Bybee 1999).

Crowder (1973) showed that the recency effect is larger than the primacy effect when subjects are asked to recall a list of stimuli. This suggested that the final-salience

effect may be due to location in the stimulus string, rather than location in the word. Bendrien (1992) has argued that word-boundary-adjacency is the determining factor because the effect does not depend on whether the first stimulus in a pair was followed by an interrupting phrase. However, the presence of a pause after the nonsense word and the interrupting phrase makes it possible to interpret the results as indicating the salience of pre-pausal, rather than word-final positions. Finally, the use of CVC stimuli confounds word-final positions, syllable-final positions, positions that involve mismatch involving the rime and the coda, and postvocalic positions.

Interestingly, the opposite effect of within-word position has been obtained in priming and word naming. Coltheart et al. (1999) found that naming the color of a visually-presented CVC word is faster when the word shares an initial phoneme/letter with the color name than when it shares a final phoneme/letter, both cases being faster than when the word and the color name do not share phonemes/letters. Kinoshita (2000) found that visually presented CVC nonsense words prime each other when they share the initial consonant but not when they share the final one. Kinoshita (2003) reviews evidence that shows that the priming effect does not occur with low-neighborhood-density words and words with irregular spelling-to-sound correspondence. From this evidence, Forster and Davis (1991) and Kinoshita (2003) have hypothesized that the priming effect occurs on a sublexical level during lexical access, specifically on the level of grapheme-phoneme correspondence rules. However, as long as similarity is assessed in a left-to-right manner and early mismatches matter more than late mismatches the effect should hold even if no grapheme-phoneme conversion is required. For instance, greater salience of initial mismatches is predicted by Marslen-Wilson's (1987,

1990/1995) Cohort Model of lexical access where all words consistent with the initial phoneme are activated and then gradually eliminated as more input becomes available. However, Radeau et al. (1995) and Slowiaczek and Hamburger (1992) find phonological priming due to final overlap to be more reliable than that based on initial overlap. Thus the salience of word beginnings may be modality- or orthography-specific.

Unlike the English studies, Tamariz (2005) found that initial matches were more salient than final matches in an AXB task in Spanish. The effect may have to do with the fact that all initial sounds were consonants while all final sounds were vowels in the study. It may also have to do with the use of visual stimuli, since visual stimuli might encourage additional left-to-right processing via grapheme-phoneme correspondence rules, especially in a language with transparent orthography like Spanish (cf. Forster and Davis 1991, Kinoshita 2003), which would increase the importance of input close to the word's beginning.

The reason initial phonemes matter more pre-lexically than post-lexically might be that during lexical access, there is a pressure to identify the word as quickly as possible, thus sounds presented first, which are usually the initial ones,<sup>21</sup> matter more. Postlexically, this pressure no longer applies and other factors, such as the level of activation in memory, become decisive. The question of whether the salience relations between initial and final segments switch after the word has been recognized, after all of its characteristics have been accessed, or once it has been perceived in its entirety needs further investigation.

Additional evidence for greater salience of initial and final sounds compared to

---

<sup>21</sup> Salasoo and Pisoni (1984) find that words can also be gated from the end and listeners can still recognize them at some point but more information is necessary for recognition.

medial ones is provided by tip-of-the-tongue (ToT) and speech error studies where the phenomenon has received the name “Bathtub effect” (Aitchison 1987). Brown and McNeil (1966) presented native English subjects with definitions of obscure English words. When subjects reported being in a ToT state, they were asked to name words that came to mind. The words were often phonologically related to the target word. When they were, the produced words and the target word shared initial and final sounds more often than medial ones (replicated by Koriat and Lieblich 1974, Rubin 1975, Browman 1978). The same result is obtained in corpus studies of malapropisms, i.e. erroneously produced words that are phonologically similar to the word appropriate in the context (Tweney et al., 1975, Fay and Cutler 1977, Hurford 1981, Aitchison and Straf 1982, see Aitchison 1987 for a review). Initial sounds are generally found to matter more than final ones. However, as Yoon and Derwing (2001) point out, there is less variation in the onset in English than in the coda, hence, more consonant sharing is expected in the word-initial position than in word-final position even if words were chosen at random.

All of the studies reviewed above agree that medial segments are less salient than segments adjacent to word boundaries. The effect is not even specific to verbal stimuli because Kidd and Watson (1992) found that sequences of tones were easier to discriminate (in terms of both speed and accuracy) for English speakers if they differed in a final or initial tone than when they differed in a medial tone. Nonetheless, this robust finding has not been incorporated into metrics of phonological similarity used in studying morphological productivity except by Albright and Hayes (2003). As noted before, in languages that have both prefixes and suffixes, the metric can be generalized to predict that word-medial phonemes are less salient than word-final and word-initial ones: when

similarity is assessed in a non-morphological task, the comparison process may be assumed to proceed from both ends of the stem.

An issue for future research is whether sound similarity in nonce probe tasks and base-derived pairing naturalness ratings behaves like sound similarity in prelexical tasks, as in postlexical tasks, or displays a third pattern. The latter is assumed by Albright and Hayes (2003). However, it is possible that domain-general similarity-assessment mechanisms can account for the sound similarity relations used in tasks measuring morphological productivity.

### **3.2.1.3. Length of match and mismatch**

Several studies investigated whether the length of both the match and the mismatch is important. Vitz and Winkler (1973) used sound similarity judgments with uni-, bi-, and tri-syllabic English words presented auditorily to native English speakers. They then used the proportion of number of mismatched segments to total number of segments in the words as a predictor of the obtained sound similarity judgments. The correlations were consistently highly significant (with the absolute value of  $r$  above 0.8), indicating that the measure is an excellent predictor of perceived similarity. However, the study did not compare the measure to measures that used only the number of matched or only the number of mismatched segments.

An explicit comparison was made in Fallon and Robinson (1987) and Kidd and Watson (1992). Kidd and Watson (1992) presented native English speakers with pairs of non-linguistic tone sequences and asked them to decide whether the sequences in a pair were different or identical. They systematically varied the duration of the match, the

duration of the mismatch and the proportion of the duration that the mismatch composed out of the total duration of the sequences. Subjects were both faster and more accurate at identifying stimuli that contained a mismatch that composed a large part of the stimulus's total duration compared to stimuli that contained a mismatch that took up a small proportion of the total duration of the stimulus when the absolute duration of the mismatch was held constant. Furthermore, when absolute duration of the mismatch was manipulated, its effect was very small compared to the effect of how much of the whole stimulus was mismatched. The effect is not limited to sequences of tones because Fallon and Robinson (1987) found that the detectability of a stretch of "frozen" Gaussian noise embedded in a longer context of random noise depends on the proportion of the total duration of the stimulus occupied by the frozen noise. However, the findings still need to be replicated with linguistic stimuli.

A step in this direction has been made recently by Peresotti and Grainger (1999) and Schiller (1998, 2004). Schiller (1998) found that the amount of priming in a word or picture naming task between a visually presented syllable prime followed by a mask, e.g. *baXXXX*, and a visually presented word target in Dutch depends on the number of shared phonemes between the prime and the target. Schiller (2004) primed Dutch words with other words, non-words, and possible Dutch word beginnings followed by a mask. For a given target, he found more priming with word-beginning primes, which contained a match but no mismatch than with real or nonsense words that had a match of the same length as word-beginnings but also contained a mismatch, e.g. *par###-parcel* vs. *pardon-parcel*. Length of mismatch was not manipulated.

While suggestive, the results do not necessarily apply to post-access global

similarity comparisons because the decrease in priming due to mismatch could be due to lateral inhibition (Rumelhart and McClelland 1981) whereby competing words inhibit each other during lexical access. The studies reported in the next section manipulate both number of matched and number of mismatched segments in acoustic linguistic stimuli in an SSJ task. In addition, Peresotti and Grainger (1999) found no difference between orthographically related words and the words' consonants in the correct order with blanks inserted for vowels used as primes in English.

The lengths of both match and mismatch have been incorporated into similarity metrics developed by Brew and McKelvie (1996), Ng (1999), Kondrak (2000), Mueller et al. (2003), and Tamariz (2005). Unfortunately, they have not been taken up in morphological investigations. Albright and Hayes (2003) do not take into account the length of either match or mismatch, focusing solely on how far from the morpheme boundary between the stem and the competing affixes the nearest mismatched segment is. Prasada and Pinker (1993) take into account only number of mismatched phonemes. Stockall (2004) takes into account only number of matched ones.

The differences between the ways of counting similarity can be illustrated with the stimuli from Stockall (2004). Stockall compares the amount of priming between present tense and past tense of English irregular verbs that differ by many vs. few segments and those that do not share morphemes but are similar both orthographically and semantically. No differences in priming were found between the two morphologically related conditions but phonologically/orthographically and semantically but not morphologically related stimuli showed no priming. In Table 3.1, we can observe that counting only number of matched segments, as Stockall does, makes formally related

stimuli that are not morphologically related appear to be much more formally similar than morphologically related stimuli. Averaging the positions of mismatches nearest to each of the stem boundaries gives the same result. On the other hand, counting the number of mismatched segments makes formally but not morphologically related stimuli appear far less related formally, while the number of matched segments/total number of segments ratio shows that the stimuli are about as formally similar as low-overlap morphologically related items.

**Table 3.1. Matched and mismatched phonemes and their locations in stimuli classified as having high or low phonological overlap in Stockall (2004)**

‘+overlap’					‘-overlap’				
stimuli	shared	mismatched	match/ total ratio	distance to mismatch	stimuli	shared	mismatched	match/ total ratio	distance to mismatch
bit-bite	2	1	0.67	2	bound-bind	3	1	0.75	2.5
bled-bleed	3	1	0.75	2.5	broke-break	3	1	0.75	2.5
chose-choose	2	1	0.67	2	brought-bring	2	2	0.5	2
came-come	2	1	0.67	2	bought-buy	1	2	0.33	1.5
dealt-deal	2	2	0.5	1.5	caught-catch	1	2	0.33	1.5
dug-dig	2	1	0.67	2	did-do	1	2	0.33	1.5
drew-draw	2	1	0.67	2	fought-fight	2	1	0.67	2
drove-drive	3	1	0.75	2.5	found-find	3	1	0.75	2.5
fed-feed	2	1	0.67	2	froze-freeze	3	1	0.75	2.5
gave-give	2	1	0.67	2	went-go	0	4	0	1
grew-grow	2	1	0.67	2	lay-lie	2	1	0.67	1.5
hung-hang	2	1	0.67	2	lit-light	2	1	0.67	2
heard-hear	2	2	0.5	1.5	paid-pay	3	1	0.75	1.5
held-hold	3	1	0.75	2.5	said-say	2	1	0.67	1.5
met-meet	2	1	0.67	2	sought-seek	1	2	0.33	1.5
rang-ring	2	1	0.67	2	sold-sell	2	2	0.5	1.5
ran-run	2	1	0.67	2	slew-slay	2	2	0.5	2
sent-send	3	1	0.75	1.5	spoke-speak	3	1	0.75	2.5
shot-shoot	2	1	0.67	2	stood-stand	3	2	0.6	2.5
sang-sing	2	1	0.67	2	stole-steal	3	1	0.75	2.5
sank-sink	3	1	0.75	2.5	struck-strike	4	1	0.8	3
sat-sit	2	1	0.67	2	swore-swear	3	1	0.75	2.5
slid-slide	3	1	0.75	2.5	took-take	2	1	0.67	2
spat-spit	3	1	0.75	2.5	taught-teach	1	2	0.33	1.5
swung-swing	3	1	0.75	2.5	told-tell	2	2	0.5	1.5
woke-wake	2	1	0.67	2	thought-think	2	3	0.4	1.5

**Table 3.1. (continued)**

‘+overlap’					‘-overlap’				
stimuli	shared	mismatched	match/ total ratio	distance to mismatch	stimuli	shared	mismatched	match/ total ratio	distance to mismatch
wrote-write	2	1	0.67	2	wove-weave	2	1	0.67	2
<b>Total</b>	<b>2.3</b>	<b>1.1</b>	<b>0.68</b>	<b>2.1</b>	<b>Total</b>	<b>2.1</b>	<b>1.6</b>	<b>0.55</b>	<b>1.9</b>
+overlap –morphology									
blossom- bloom	3	3	0.5	2.5	scald-scorch	2	3	0.4	2
boil-broil	3	1	0.75	2	scrape-scratch	3	2	0.6	2.5
burst-bust	3	2	0.6	2.5	screech-scream	4	1	0.8	2.5
converge- merge	3	4	0.43	2	shimmer- glimmer	5	2	0.71	3.5
crinkle- wrinkle	5	1	0.83	2.5	shrivel-shrink	3	3	0.5	2
crumple- rumple	5	1	0.83	2.5	slim-trim	2	2	0.5	2
flip-flop	3	1	0.75	2.5	strain-sprain	4	1	0.8	3
ghost-ghoul	1	3	0.25	1.5	squish-squash	4	1	0.8	3
gloom-glum	3	1	0.75	2.5	drop-drip	3	1	0.75	2.5
mangle-tangle	4	1	0.8	2.5	clash-clang	3	1	0.75	2
pat-pet	2	1	0.67	2	<b>Total</b>	<b>3.2</b>	<b>1.8</b>	<b>0.55</b>	<b>2.4</b>
plunge- plummet	3	4	0.43	2.5					

#### **3.2.1.4. Psychological reality of sublexical phonological units**

Most of the work on sound similarity has been focused on determining which sublexical phonological units hearers are sensitive to. In particular, all studies conducted so far have focused on whether hearers are sensitive to the number of matched and mismatched units of some type (issue 3a). McInish and Tikofsky (1969) found that native English speakers could distinguish non-words that differed by one segment faster and more accurately if the segments differed in two features (drawn from Miller and Nicely 1955) than if they differed in one feature. However, as McInish and Tikofsky point out, it is not clear whether the effect is purely acoustic in nature or specifically linguistic units are involved.

Derwing and Nearey (1986), and Bendrien (1992) replicated this finding with sound similarity judgments but furthermore found that if the two mismatched features were distributed over the final and the initial segments of CVC stimuli, native English speakers perceived the words to differ more than when the two features were located within a single segment. Tamariz (2005) found that in Spanish, CVCV and CVCCV words that share two segments are perceived to be more similar in sound than those that share one segment in a visually-based AXB task.

While these findings appear to support the notion of a segment, as proposed by Vitz and Winkler (1973), segment number is confounded with syllabic constituent number when consonantal mismatches in CVC syllables are examined. In addition, consonants at two word boundaries are mismatched in the two-consonant mismatch condition as opposed to a consonant at only one word boundary in the one-consonant mismatch condition. The results could also be due to the two-consonant condition containing a non-contiguous mismatch (cf. Kidd and Watson 1992, who observed a

difference in detectability of continuous and non-continuous mismatches of the same overall duration), rather than being due to number of segments. Finally, differences in temporal duration could also account for the results.

Treiman (1983), Derwing et al. (1987, 1991), Dow (1987), Treiman and Danis (1988), Derwing and Wiebe (1994), Yoon (1994), and Yoon and Derwing (2001) have all provided support for the notion of the syllabic constituent as a unit intermediate between segments and syllables.

All of the studies on English have converged on an onset-rime analysis of the syllable. For instance, Treiman (1983), using a string manipulation technique in which subjects were asked to form one CCVCC word out of two that were presented by combining the beginning of the first word with the end of the second word, found that subjects overwhelmingly preferred to combine the onset of the first word (CC) with the rime of the second word (VCC). The result was replicated with unit substitution (Derwing et al. 1987), unit counting (Dow 1987), serial recall (Treiman and Danis 1988), and sound similarity judgments (Yoon and Derwing 2001).

Radeau et al. (1995) review evidence from several phonological priming studies that shows that priming is more robust in English when the rime is shared than when the body is shared between the prime and the target. Radeau (1995) found that phonological priming did not occur in English nonsense  $C_1C_2VC_3$  words when only the coda was shared but did occur when the rime was shared. Interestingly, there was no difference in the magnitude of priming between cases when only the rime was shared and cases when the second consonant of the onset was also shared. On the other hand, Dumay et al. (2001) found more priming with matches involving the onset and the rime of a second

syllable in bisyllabic stimuli than with matches involving only the rime. These findings suggest that sharing parts of two units does not produce as much similarity as sharing one whole unit. That is, if there is some mismatch in the unit, increase in the match does not matter because the unit is still not shared. However, it may also be that the priming task is simply not sensitive enough to detect differences in similarity due to increase in overlap length. In section 3.2.3 we show that subjects are in fact aware of differences in mismatch duration even when the difference is not phonologically meaningful.

Beinert and Derwing (1993) found that, while words that differed by more segments were perceived to be less similar in Arabic, the consonantal tier also accounted for a significant amount of additional variance. Similarly, Derwing and Wiebe (1994) found that the mora accounted for additional variance in Japanese, compared to counting the number of mismatched segments.

Yoon and Derwing (2001) used a variety of techniques to compare syllable structures of English and Korean. They found that in English CVC non-words, both English monolinguals and English-Korean bilinguals find stimuli that mismatch in the nucleus and the coda to be more similar than stimuli that mismatch in the nucleus and the onset. However, the opposite is true for Korean CVC non-words for both Korean monolinguals and Korean-English bilinguals who are illiterate in Korean. Furthermore, the difference is found even for a single group of bilingual speakers. These results strongly suggest that the nucleus and the coda form a unit in English (the rime) while the onset and the nucleus form a unit in Korean (the body). Thus, a CV mismatch in English involves four units, while a CV mismatch involves only three, and the opposite is true in Korean. The results were replicated with concept formation, unit reduplication accuracy,

and serial recall (Yoon and Derwing 2001), as well as Treiman's (1983) word blending task (Derwing et al. 1991, Wiebe and Derwing 1994, Yoon 1994).

Yoon and Derwing (2001) propose an explanation for the differences in syllable structure in English and Korean. They note that there are more body types than rime types in Korean while the opposite is true in English. Therefore, the Korean child is exposed to much more body-based variability than rime-based variability and learns to attend to differences in bodies while the opposite is true for the English-learning child. An alternative explanation is that codas are somewhat predictable given the nucleus in English while nuclei are not predictable given the onset (Kessler and Treiman 1997). However this hypothesis remains to be tested for Korean. Both hypotheses need to be tested on a larger sample of languages. Additional predictions that follow from an account of syllabic constituents as emergent is that speaker/hearers should be sensitive to transitional probabilities between segments (demonstrated for English by Coleman and Pierrehumbert 1997 and Frisch et al. 2000) and/or constituent frequencies (not yet tested) and that subsyllabic structure within a language may depend on the segments involved. For instance, it is plausible that tense and lax vowels might behave differently in English, given Kessler and Treiman's (1997) account of constituent emergence: lax vowels require a coda and thus might be more likely to fuse with it than tense vowels. In chapter 7, we will provide evidence that bodies and rimes exist in Russian and thus it is plausible to hypothesize that they also exist in English but differ in their resting activation levels.

Tamariz (2005) has examined whether speakers are sensitive to syllables in judging sound similarity. She compared whether CV.CCV stimuli in Spanish were more similar to other CV.CCV stimuli than to CVC.CV stimuli. No effect was found.

However, no direct comparison was made. Rather, sharing syllabic structure was never more salient than any other type of similarity, e.g. sharing a segment. Furthermore, words that shared a consonant from each syllable (C2 and C3) were perceived to be less similar than those that shared consonants from a single syllable (C1 and C2), contrary to what counting syllables that are matched and mismatched would predict. However, this effect is confounded with salience of the initial position, which could override any effect of syllable number. Furthermore, it is not yet known whether an increase in similarity that does not result in unit identity is more or less salient within a unit than across units.

Titone and Connine (1997) presented native English speakers with CVCCVC stimuli in a priming task where the prime was always a pseudoword, while the target was either a word or a pseudoword. The prime and the target always shared all phonemes but the initial one. Primes were artificially syllabified by introducing a pause either right after the nucleus of the first syllable or before the nucleus of the second syllable. Primes syllabified after the onset of the first syllable produced more priming. Titone and Connine suggested that the findings provide support for the notions of a syllable and the onset maximization principle. While onset maximization is indeed supported, the results are also consistent with an account that does not postulate syllables and instead uses only syllabic constituents.

The effect of the number of mismatched syllables in English nonsense words where location of mismatch, number of mismatched and matched segments, and number of matched and mismatched syllabic constituents are controlled is explored in the next section.

### **3.2.1.5. Stress**

Tamariz (2005) is the only study to have explored the effect of stress on match salience. Matches in stressed vowels were more salient than matches in unstressed vowels, even when matches in stressed medial vowels and unstressed final vowels are compared. Furthermore, words that shared stress location were found to be more similar than those that did not. Again, however, no direct comparisons were made. Interesting issues still remain. For instance, we do not yet know whether consonants in stressed syllables are more salient than consonants in unstressed syllables.

### **3.2.1.6. Type of mismatch**

Steriade (2001, 2004) proposes that the acoustic perceptibility of a segment influences its salience in similarity judgments. Thus the effects of within-word location and phonological salience may be different for different types of mismatches.

Steriade (2001) analyzes regressive and progressive place assimilation in CC clusters. She hypothesizes that if /anpa/ is perceived to be more similar to /ampa/ than to /anta/, it will be realized as /ampa/. Fujimura et al. (1978), Ohala (1990), and Jun (1995) find that when a pause is preceded by transitions indicating one consonant and followed by transitions indicating another, hearers make the judgment of what the consonant is based on the CV transition, not the VC one in both English and Japanese. Furthermore, when the stimuli were played backwards, CV transitions, which were the VC transitions in the previous study, were still decisive (Fujimura et al. 1978). These results indicate that in VC<sub>1</sub>C<sub>2</sub>V, place assimilation of C<sub>1</sub> to C<sub>2</sub> should create a smaller perceptual mismatch than assimilation of C<sub>2</sub> to C<sub>1</sub> (Steriade 2001). On the other hand, retroflexion

is primarily determined by VC transitions (Ladefoged and Maddieson 1986). Thus, assimilation in retroflexion should affect C<sub>2</sub>. This is precisely what has been found by Steriade (2001) in a cross-linguistic study.

Steriade (2004) conducted a corpus study of imperfect rhymes in Romanian poetry. In Romanian poetry, voicing mismatches in oral stops are more likely to occur in N\_# than in N\_V where it is more likely than in V\_# and V\_V for every one of four poets studied<sup>22</sup>. The result is explained by an acoustic difference: Steriade and Zhang (2001), summarized in Steriade (2004), measured differences in voicing duration/total duration ratios between voiced and voiceless stops in postnasal and postvocalic environments produced by two Romanian speakers and found that the difference in VOT between voiced and voiceless stops was smaller in postnasal contexts. Thus, the shorter the temporal duration of a mismatch, the less salient it is perceived to be.

Steriade (2004) argues that loan adaptation may also provide evidence of similarity judgments, e.g. in deciding to simplify a CVC<sub>1</sub>C<sub>2</sub> input as CVC<sub>2</sub> the speaker judges C<sub>1</sub> to be less perceptible than C<sub>2</sub> and thus that C<sub>1</sub>C<sub>2</sub> is more similar to C<sub>2</sub> than to C<sub>1</sub>, and that C<sub>1</sub>C<sub>2</sub> is more similar to C<sub>2</sub> than to C<sub>1</sub>VC<sub>2</sub>. However, no evidence regarding the similarity judgments made by the same speakers who use a certain way of loanword adaptation is provided.

### **3.2.1.7. Summary**

To summarize, prior research indicates that

- 1) initial and final segments are more salient than medial segments,
- 2) stressed vowels are more salient than unstressed vowels,

---

<sup>22</sup> N-nasal consonant, V-vowel

- 3) both the length of the match and the length of the mismatch matter,
- 4) number of matched and mismatched syllabic constituents matters, although where the syllable is divided differs from language to language,
- 5) acoustic salience of mismatch matters and thus different types of mismatches can be affected by location and surrounding context differently.

In the following two sections, I use sound similarity judgments to provide evidence that

- 1) number of matched and mismatched syllabic constituents matters but number of segments does not, contrary to claims in the literature, e.g. Vitz and Winkler (1973), Bendrien 1992, Yoon and Derwing (2001); rather, the effect of number of mismatched segments is due to temporal duration, just as shown by Kidd and Watson (1992) for number of matched and mismatched tones vs. temporal duration in nonlinguistic stimuli, cf. also Steriade (2004);
- 2) both duration of the matched and duration of the mismatched matter;
- 3) final consonants are much more salient than initial consonants, which are much more salient than medial consonants that are part of the initial constituent, which are a little more salient than medial consonants that are part of a medial constituent.

### **3.2.2. Experiment A: Number of mismatched syllabic constituents, location of mismatch, and number of matched segments<sup>23</sup>**

In pilot work with English speakers, using the AXB task with monosyllabic stimuli (Kapatsinski 2004), it was found that words that differ by two non-contiguous consonants

---

<sup>23</sup> This section has been presented at VIII Encuentro Internacional de Linguística en el Noroeste and will appear in the conference proceedings (Kapatsinski 2004e).

are perceived to differ more than words that differ by two contiguous consonants. This section presents evidence that this effect is due to the fact that words mismatched by two non-contiguous segments differ by two syllabic constituents (an onset and a coda) while words differing by two contiguous segments differ by only one syllabic constituent. Bisyllabic words are used to unconfound number of mismatched syllabic constituents and contiguity of mismatched segments. While the number of mismatched syllabic constituents is shown to have a reliable effect on the perceived similarity of bisyllabic words in the SSJ<sup>24</sup> task, contiguity does not have a reliable effect. Neither does number of mismatched syllables.

### **3.2.2.1. Method**

Thirty-six native English speakers were presented auditorily with 116 pairs of nonsense, phonotactically legal bisyllabic words. Subjects were asked to rate words in each pair on how similar they sounded to each other on a 21-point scale ranging from 1 “as different as two words could possibly be” to 21 “the same; completely identical”. All words were produced by the same native English speaker. A pair was presented twice with two-second pauses between words and a seven-second pause for the subject to respond. Words in each pair differed by two consonants. The mismatches were of four types: 1) where one word has /k/, the other has /n/, and where the first word has /s/, the other has /t/; 2) where one word has /n/, the other has /p/, and where the first word has /s/, the other has /t/; and 3) where one word has /k/, the other has /p/, and where the first word has /s/, the other has /t/. The reason these alternations were used is that the clusters produced by combining these phonemes (/ks/, /nt/, /pt/, /ns/, /ps/, /nt/) were attested with a type

---

<sup>24</sup> Sound Similarity Judgment

frequency greater than 3 in all positions in bisyllabic English words in the Switchboard Corpus (Godfrey et al. 1992). and the resulting alternants differed by more than 1 feature. Type of mismatch was controlled such that a quarter of stimuli in each experimental condition contained alternations of type 1, half – alternations of type 2 and a quarter – alternations of type 3. Half of the stimuli in each experimental condition had stress on the first syllable and half on the second syllable. Listener identity and number of matched segments were entered as independent variables together with the experimental variable of interest into an ANOVA and only results that remained significant in competition with listener identity and number of matched segments are reported here. Since 8 ANOVA's were run over intersecting sets of stimuli, the .05 significance level with the Boferroni correction in this study is at .00625. All effects reported here were found to be significant at least at the .0005 level. The additional potentially confounding factor of number of mismatched consonants that are part of a consonant cluster was independently tested and found to have no effect. The stimuli are listed in Table 3.2. They were presented in an order that was random except for ensuring that two stimuli from the same condition never appeared within five stimuli of each other. Subjects were tested alone with stimuli presented over headphones. No noise was added.

**Table 3.2. Stimuli used in Experiment A**

Mismatch type	Stimuli
Contiguous; Medial (late)-Final; One syllable; One constituent;	ɪpræks - ɪprænt; rɪkɫent - rɪkɫeps; əbrept - əbrens; ækɫɪpt - ækɫɪks; bɛrnənt - bɛrnəks; pansəns - pansəpt; reɪnənt - reɪnəps; partəks - partəpt; pɪgəuks - pɪgəunt; brɪntens - brɪntept; krɪlənt - krɪləps; arfeks - arfept; tunəks - tunənt; wɛrnəns - wɛrnəpt; arlənt - arləps; kɛrtɪks - kɛrtɪpt; rɪmouks - rɪmount; hoɪlənt - hoɪləps; kraɪsəns - kraɪsəpt; pɪrhouks - pɪrhoupt
Contiguous; Medial-Medial; Two syllables; Two constituents;	fɪksɛrt - fɪntɛrt; brænsɛst - bræptɛst; klantənt - klapsənt; pleptɪks - pleksɪks; ɪntɛp - ɪksɛmp; wɪnsɛɪt - wɪptɛɪt; ɪntɪn - ɪpsɪn; gæptɛr - gæksɛr; bruntɪ - bruksɪ; kənsɛr - kəptɛr; krontɪst - kropsɪst; snɪptəl - snɪksəl; lɪksɛnt - lɪntɛnt; ɪnsaund - ɪptaund; lɪntaɪd - lɪpsaɪd; lɪksɛr - lɪptɛr; trɪksɛɪn - trɪntɛɪn; əntɛɪd - əpsɛɪd; tauksɛr - tauptɛr; krænsɛr - kræptɛr
Non-contiguous; Medial (late)-Final; One syllable; Two constituents; 0 C's part of clusters	pɪrkəus - pɪrɪnaut; baɪnəs - baɪpət; kərnɛɪt - kərpɛɪs; rækəs - ræpət; wəkəs - wonət; smɪnəɪs - smɪpəɪt; wonət - wopəs; kwəkəs - kwəpət; vəkəns - vənənt; senəs - sepət; kənɛɪt - kəpɛɪs; lɪkəɪs - lɪpəɪt; məkəs - monət; wonəs - wopət; dɛrnənt - dɛrpəns; bləkɪs - bləpɪt; dɪkəs - dɪnət; bɪnɛɪs - bɪpɛɪt; drɪnaut - drɪpəus; sɪkəus - sɪpəut
Non-contiguous; Initial-Medial; One syllable; Two constituents; 0 C's part of clusters	nausɛr - pautɛr; nartɪ - parsɪ; krouzən - proutən; kɪsɛnt - nɪtɛnt; nəsɛɪt - pətɛɪt; notəɪzd - posəɪzd; kɪsəud - pɪtəud; kɪsəɪn - pɪtəɪn; kəusɛrt - nautɛrt; nɪsənt - pɪtənt; notərd - posərd; kəsən - potən; kəsɛɪ - nətɛɪ; nəsəl - pətəl; nətəɪt - pəsəɪt; kɪsəund - pɪtəund; kəsərn - notərn; nɪsul - pɪtul; natəl - pasəl; kɪsouz - pɪtəuz
Non-contiguous; Initial-Final; One syllable; Two constituents; 0 C's part of clusters	kəɪpərs - nəɪpərt; nɛɪkəs - pɛɪkət; nostrɪt - postɪs; kəɪəs - pəɪtɛt; kəprəɪs - nəprəɪt; nəmbɪt - pəmbɪs; kəros - pərot; nərdəut - pərdəus; kəmbəs - nəmbət; nukləs - puklət; nərmɪt - pərmɪs; kəɪnəs - pəɪnət; kəmfərs - nəmfərt; nədrəɪs - pədrəɪt; norɪət - pəɪtɛs; kərmɪəs - pərmɪt; kəpɛs - nəpɛt; noumɪs - poumɪt; nərvɛɪt - pərvɛɪs; kɪvəus - pɪvəut

**Table 3.2 (continued).**

Mismatch type	Stimuli
Non-contiguous; Medial (late)-Final; One syllable; Two constituents; 1 C part of cluster	krikəns - krinənt; wounəns – woupənt; kwɪnɪft - kwɪpɪfs; soukɪns - soupɪnt; wəkæps - wənæpt; rɪnuks - rɪpukt; kunekt – kupeks; dʒɪkafs – dʒɪpɑft
Non-contiguous; Medial (late)-Final; Two syllables; Two constituents; 1 C part of cluster	tuksɪs - tunsɪt; kontərs - koptərt; vonsərt - vopsərs; læksəns - læpsənt; rɪkses - rɪnset; kənsult - kəpsuls; dʒʊntet - dʒʊptes; ɪksos – ɪpsot
Non-contiguous; Initial-Medial; Two syllables; Two constituents; 1 C part of cluster	kaunsər - nauntər; næpsənt - pæptənt; nuftɪəl - pufsiəl; kospeɪt - postɪt; kəpsər - næptər; nəsats - pətats; nəktərn - pæksərn; kɪnsoud – pɪntoud
Non-contiguous; Medial (early)-Final; Two syllables; Two constituents; 1 C part of cluster	skopləs – snoplət; snorkɪs - sporkɪt; snulɪt - spulɪs; skantrəs - spantrət; skɪlars - snɪlart; snərouz - spərou; snəkɪt - spəkɪs; skərleɪs – spərleɪt
Non-contiguous; Medial-Medial; Two syllables; Two constituents; 1 C part of cluster	skɛɪsənt - sneɪtənt; snusər - sputər; snartəl - sparsəl; skousədʒ – spoutədʒ; skuɪzɪz - snuɪzɪz; snəsaul - spətaul; snɪtərn - spɪsərn; skərsoud – spərtoud
Non-contiguous; Medial-Medial; Two syllables; Two constituents; 2 C's part of clusters	skoupsər – snouptər; snɪksərn - spɪktərn; snæftən - spæfsən; skulsənd - spultənd; skɪpsɪt - snɪptɪt; snəfsoɪn - spəftoɪn; snɪntərn - spɪnsərn; skəlsoud – spəltoud
Non-contiguous; Medial (late)-Final; Two syllables; Two constituents; 2 C's part of clusters	fæksəns - fænsənt; pauntəns - pauptənt; wɪnsənt - wɪpsəns; trɪksəps - trɪpsəpt; bɪksɛps - bɪnsept; wɪnsauns - wɪpsaunt; fləntent - fləptens; drəksouks – drəpsoukt
Non-contiguous; Medial (early)-Final; Two syllables; Two constituents; 2 C's part of clusters	skæmbɛns - snæmbɛnt; snoləks - spoləkt; snarwənt - sparwəns; skərɪlɪfs - spərɪlɪft; skərpapɪs - snərpapɪt; snəmækt - spəmæks; snɪkroups - spɪkroupt; skɪrauns – spɪraunt

### 3.2.2.2. Number of mismatched syllabic constituents and mismatch location

To test the idea that the contiguity effect is due to the number of mismatched syllabic constituents, we made two comparisons: 1) words that differed by two contiguous segments, one of which was the final consonant, in a single syllabic constituent were compared with words that differed by two non-contiguous segments, one of which was the final consonant, in two syllabic constituents, and 2) words that differed by two non-contiguous segments, neither of which was the final consonant, in two syllabic constituents were compared with words that differed by two contiguous segments, neither of which was the word-final consonant, in two syllabic constituents. The reason the comparisons were made this way is that mismatches involving the final consonant were found to be extremely salient with all types of mismatches tested as shown in Table 3.3.

**Table 3.3. Mismatch salience as a function of location in bisyllabic words**

Location of mismatched C's	Medial+Medial skɛɪsənt snɛɪtənt	Medial+Initial kosəɾn notəɾn	Medial+Final fæksəns fænsənt	Initial+Final kaɪpərs naɪpərt
Mean score of perceived similarity (the higher the score, the more similar the words)	9.441	9.266	7.762	6.209
Significance	n.s.		p<0.0005	
		p<0.0005		

Thus comparing mismatches in two contiguous segments in the final coda with mismatches in two word-internal contiguous segments belonging to different constituents pits the salience of mismatches in the final position against the salience of mismatches involving syllabic constituents resulting in observing no difference between these two conditions (Table 3.4) but the numerical tendency suggests that number of mismatched syllabic constituents exerts a slightly stronger influence. The perceived similarity of words in which both factors agree is significantly different from both cases of conflict,

indicating that both factors influence the subjects' behavior.

**Table 3.4. Mismatch in two contiguous segments in a single constituent that include the word-final segment versus mismatch in two medial contiguous segments in two constituents**

Condition	1 syllabic constituent Medial + Final ɪpræks - ɪprænt	2 syllabic constituents Medial + Medial fɪksɜrt – fɪntɜrt	2 syllabic constituents Medial + Final pɪrɪkaus - pɪrɪnaut
Mean Score	10.364	9.907	7.762
Significance	n.s.		p<0.0005 (vs. either other condition)

The hypothesis that number of mismatched syllabic constituents influences the perceived phonological/phonetic similarity of words is confirmed by the data in Tables 3.5-3.6, which show that while words that are mismatched by two syllabic constituents are reliably perceived to be more different than those that are mismatched by one syllabic constituent, contiguity of mismatched segments does not have a significant effect.

**Table 3.5. Mismatch in two contiguous segments in one constituent vs. two non-contiguous segments in two constituents where one of the segments is word-final in both cases**

Condition	1 syllabic constituent Contiguous ɪpræks - ɪprænt	2 syllabic constituents Non-contiguous pɪrɪkaus - pɪrɪnaut
Mean Score (p<0.0005)	10.364	7.762

**Table 3.6. Mismatch in two contiguous segments in two constituents vs. two non-contiguous segments in two constituents where none of the segments are word-final in both cases**

Condition	2 syllabic constituents Contiguous fɪksɜrt – fɪntɜrt	2 syllabic constituents Non-contiguous kæpsɜrn-næptɜrn
Mean Score (n.s.)	9.907	9.021

Table 3.7 shows that the number of mismatched syllables does not have any effect by comparing words mismatched in the onset and coda of the first syllable with words mismatched in the onset of the first syllable and the onset of the second syllable.

**Table 3.7. Mismatch in two non-contiguous segments in two syllables vs. in one syllable**

Condition	1 syllable krikəns - krinənt	2 syllables tuksɪs - tunstɪt
Mean Score (n.s.)	7.776	7.786

Table 3.8 shows that while mismatches beginning early in the middle of the word are reliably more salient than those beginning late in the middle of the word as predicted by the Cohort Model of lexical access (Marslen-Wilson 1990/1995), the difference is very small and thus cannot explain the difference between words mismatched in one vs. two syllabic constituents. The difference could also be explained as a word-boundary effect on the syllabic constituent level, since mismatches beginning early involve word-initial syllabic constituents while mismatches beginning late involve word-medial constituents.

**Table 3.8. Mismatches beginning in a word-medial consonant close to word beginning vs. those beginning in a medial consonant close to word end**

Condition	Mismatch begins early skæmbəns - snæmbənt	Mismatch begins late fæksəns - fænsənt
Mean Score (p<0.001)	7.743	7.803

The results clearly indicate that syllabic constituents are psychologically real for native English speakers and play a stronger role in determining how similar two words sound than syllables.

### 3.2.2.3. Number of matched segments

When tested in the same ANOVA as subject identity and a variable incorporating number of mismatched syllables, number of mismatched syllabic constituents, location of mismatch, and number of mismatched consonants that are part of consonant clusters the effect of the number of matched segments was significant at the .0005 level, such that the

more segments two words share, the more similar they are (table 3.9).

**Table 3.9. Effect of number of matched segments<sup>25</sup>**

Number of matched segments	3 wokəs wonət	4 nisənt pitənt	5 snorkɪs sporkɪt	6 skulsənd spultənd	7 skərpəɪps snərpəɪpt
Mean Score (p<0.0005)	7.79	8.28	8.87	10.17	11.87

### 3.2.3. Experiment B: Segments and milliseconds<sup>26</sup>

#### 3.2.3.1. Method

Sixty-eight native English speakers were presented with pairs of nonsense monosyllables that differed by either one segment or two contiguous segments. The mismatch always involved the word-final consonants. All pairs contained words that differed in place, voice, manner, and nasality (e.g. /nt/-/bz/, /s/-/m/). All stimuli were produced by the same native English speaker and then altered in duration. There were two-second pauses between words and eight seconds for subject response. Each pair was repeated twice. For half of the subjects the stimuli that differed by two segments and those that differed by only one differed by stretches of speech of the same temporal duration (15 ms). For the other group, the differences in duration between mismatched stretches of speech containing two segments and those containing one were artificially increased so that words that differed by one segment differed by 7 ms of sound while those that contained two differed by 30 ms. In each group, 7 subjects were asked to write the stimuli down. Only stimuli that were spelled in the same way by all 14 control subjects were included in the experiment to ensure that the short and long versions of the stimuli were

<sup>25</sup> For the purposes of this analysis, diphthongs were counted as two segments. Postvocalic /r/ was counted as a separate segment.

<sup>26</sup> This section has been presented at VIII Encuentro Internacional de Linguística en el Noroeste and CLS 41 (Kapatsinski 2005c). It will appear in the Encuentro proceedings Kapatsinski (2004e).

phonologically identical and fully perceptible. The resulting stimuli were matched so that for each pair that differed in two segments there was a pair that differed in one segment with the same onset and nucleus and vice versa. Thus length of matched stretches of speech and number of matched segments were controlled. The stimuli are presented in Table 3.10. Subjects were tested alone with stimuli presented over headphones. No noise was added. The data were analyzed using two ANOVA's where number of mismatched segments competed against listener identity.

**Table 3.10. Stimuli used in experiment B**

Number of mismatched segments →	1	2
	kw <b>ɪ</b> n-kw <b>ɪ</b> f	kw <b>ɪ</b> n-kw <b>ɪ</b> ks
	<b>g</b> ren- <b>g</b> reps	<b>g</b> ren- <b>g</b> ref
	h <b>æ</b> nt-h <b>æ</b> sp	h <b>æ</b> nt-h <b>æ</b> ft
	klent-kleps	klem-kles

### 3.2.3.2. Results and Discussion

Tables 3.11-3.12 show that words that differed by two segments were only perceived to differ more than words that differed by one segment if they differed by a stretch of speech of greater temporal duration than words that differed by one segment. Moreover, the same allophone of the same phoneme is given more weight if it is phonetically longer ( $p < 0.0005$ ).

**Table 3.11. Words differing by one segment vs. two contiguous segments: No differences in temporal duration of mismatched stretches of speech**

Condition	1-segment difference kw <b>ɪ</b> n-kw <b>ɪ</b> f	2-segment difference kw <b>ɪ</b> n-kw <b>ɪ</b> ks
Mean Score (n.s.)	7.89	8.95
Mean temporal duration of mismatched stretches of speech	15.2 ms	15.1 ms

**Table 3.12. Words differing by one segment vs. two contiguous segments: Increased differences in temporal duration of mismatched stretches of speech**

Condition	1-segment difference kwIn-kwIf	2-segment difference kwInt-kwIks
Mean Score (p<0.0005)	11.42	8.11
Mean temporal duration of mismatched stretches of speech	6.9 ms	30.3 ms

The results reported in this section indicate that sound similarity judgments are made based on detailed phonetic information about the particular tokens of the words that are being compared, including the phonologically redundant duration of parts of those words. This finding provides support for exemplar-based models of the mental lexicon (e.g. Hintzman 1986) in which every exposure to a stimulus creates a new token representation. This contradicts distributed connectionist models, which include a fixed set of nodes over which all words are represented (e.g. Rumelhart and McClelland 1986), since tokens are not kept distinct in such models or any models that discard perceptual detail and normalize the stimulus to a fixed representation (cf. Johnson 1997).

The results say nothing of the length of time for which the tokens are stored. More work is also needed to determine whether the results hold across mismatch types and to confirm the results with a larger stimulus sample. In addition, it would be interesting to see if the temporal duration of matched stretches of speech is a better predictor of perceived similarity than number of matched segments.<sup>27</sup>

### **3.2.4. General discussion: Implications for the study of morphological productivity**

The findings of this section show that both number of matched segments and number of

---

<sup>27</sup> While the stimuli may have sounded rather unusual, they would have sounded unusual in all conditions because a mismatch of a given number of segments was shortened by the same number of milliseconds in one condition by which it was lengthened in the other condition.

mismatched segments must be taken into account in determining how phonologically similar two words are (see Fallon and Robinson 1987, Kidd and Watson 1992, Schiller 1998, 2004). This has important methodological implications for studies of neighborhood effects in priming and morphological productivity since neighbor similarity is usually assessed only by measuring the number of mismatched segments. Even the definition of ‘neighborhood’ has been adopted from orthography where it was never tested either. Thus the standard definition for ‘neighbor’ in orthography is ‘any word that can be created by changing one letter of the stimulus and preserving letter positions’ (e.g. Coltheart et al. 1977, Carreiras et al. 1997) and in phonology ‘any word that can be created by changing (or deleting) one phoneme of the stimulus (or inserting a phoneme) and preserving phoneme positions’ (e.g. Luce and Pisoni 1998, Ziegler et al. 2003). In addition, number of shared phonemes is often not controlled in studies of morphological priming that aim to test whether it is independent of phonological similarity (e.g. Stockall 2004). Furthermore, the results presented here show that number of matched and mismatched syllabic constituents, location of mismatch relative to word boundaries and the temporal duration of both match and mismatch matter and must be taken into account by any sound similarity metric. Finally, one needs to control the actual acoustics of the stimuli, not just impose restrictions on their phonological representations.

### ***3.3. The similarity metrics used in this study***

Both metrics used in the following experiments use the fixed radius method. While for both studies the UNOS was based on the Halle (1995) feature, the definition of UNOS differed greatly.

In experiment I, Feature changes implied by more specific changes were not counted, e.g. the change from [d] to [n] involves changing the features [nasal] and [sonorant] but the change in the feature [sonorant] was not counted. In experiments II and III feature changes implied by more specific changes were counted because of the finding, reported by McInish and Tikofsky (1969) that number of mismatched features is correlated with perceived similarity of English CVC stimuli that differ by one consonant. The change from one value to another of the non-binary feature [place] was always counted as 1 UNOS in both experiments, regardless of the distance between values in articulatory space.

Only monosyllabic roots were used in all experiments. In experiment I, words in the neighborhood were ‘derived’ from the stimulus by the addition or deletion of any number of consonants to the beginning or end (but not the middle or both) and any number of deletions and insertions within a single syllabic constituent was counted as 1 UNOS. This method is consistent with the notion of syllabic constituents but does not take into account differences in temporal duration between different mismatches (found to matter by Vitz and Winkler 1973, Bendrien 1992, Kidd and Watson 1992, Kapatsinski 2004e, 2005c, this thesis, Steriade 2004, and Tamariz 2005). Therefore, in experiments II and III a mismatch in X UNOS distributed over two consonants differed in 1.1X UNOS.

The coefficient of 1.1 was empirically derived as the average of the ratios in the two conditions of Experiment B. No other study provides relevant data because Vitz and Winkler (1973) and Bendrien (1992) confound number of mismatched segments and number of mismatched syllabic constituents, while Kapatsinski 2004, Steriade 2004, and Tamariz 2005 do not allow us to estimate the magnitude of the difference because of

examining binary decisions. Thus, if for all speakers words that differ by two consonants are 5% more similar than words that differ by one consonant, 100% of them may still choose words that differ by two consonants as being more similar to the base in an AXB task. An addition or deletion of a consonant was counted as 1 UNOS in experiment II. Only monosyllabic models were considered in both Experiments.

Substitutions of allophones of the same phoneme<sup>28</sup> were counted as changes by  $\frac{1}{2}$  UNOS in all experiments, as were substitutions of one phoneme for another in environments where they would be realized in the same way phonetically in all experiments. For example, voiced obstruents devoice at the ends of words in Russian, hence the phonemic contrast between /t/ and /d/ is neutralized in this position and the distance between /kot/ and /kod/ is half a unit. This assumption awaits empirical testing.

In experiment II, a mismatch in X UNOS in the final position was counted as 1.2X UNOS, the coefficient being derived from experiment A. Because experiment A used bisyllabic stimuli, the relevance of the coefficient to CVC stimuli used in experiments II and III is not apparent. However, the same coefficient is apparent in Bendrien's (1992) results with CVC stimuli. Finally, in experiment II, a mismatch in X UNOS that was distributed across the onset and the rhyme was counted as 1.4X UNOS, the coefficient being derived from experiment A. Either the penalty for cross-segment distribution or for cross-constituent distribution was applied to a given pair of stimuli.

The radius in all experiments was 4 UNOS (based on Connine et al. 1993, who found that monosyllabic words that differ by 4 features do not prime each other in phonological priming). This means that the neighborhood was narrowed in experiments II and III than in experiment I. Chapter 8 provides evidence for a small neighborhood

---

<sup>28</sup> Resulting from substituting adjacent phonemes

radius. All stimuli shared at least one segment. Number of segments was reflected in the temporal durations of match and mismatch.

A control set of words that are not similar to any of the existing words was derived for each experiment. Such a control set is necessary to evaluate the productivity of various patterns and its dependence on similarity (Clahsen 1999). According to the DMM, the pattern used most in this set of stimuli is the default pattern. With the FIRM, this is a set of words that do not have any neighbors within the radius or whose neighbors from different classes are equal in number and equidistant from the stimulus.

### ***3.4. Future directions***

In this chapter, we have seen evidence that location within the word, number of matched and mismatched syllabic constituents, and temporal duration of the match and mismatch influence how similar two words sound. We have also developed a new method of controlling for phonological similarity and a similarity metric that incorporates the basic findings in the literature.

However, several gaps in our knowledge at this point are apparent. One issue is cross-linguistic variation. Most of the work on sound similarity has been done on English and Yoon and Derwing's (2001) work on Korean, Beinert and Derwing's (1993) work on Arabic, and Derwing and Wiebe's (1994) work on Japanese have made it clear that linguistic experience influences similarity judgments. Thus for working on Russian morphology, we need to conduct sound similarity judgment experiments with native Russian speakers to develop a sound similarity metric for Russian.

Another issue is that we do not know how word length impacts the importance of determinants of sound similarity. For instance, it is likely that location in relation to word boundaries is a stronger factor in long words than in short ones. It might even be the case that relative salience of initial and final sounds interacts with word length because of differences in short-term memory demands and the likelihood of accessing the word in a left-to-right manner.

Furthermore, we do not know if there are nonlinearities in how match/mismatch ratio influences sound similarity. For instance, stimuli that contain a short mismatch embedded in a long match might draw attention to properties (e.g. length) of the mismatch, and away from properties of the match while the opposite may be true with short matches embedded in long mismatches. A related issue is whether a difference between two lengths produces the same difference in sound similarity, regardless of what the lengths are. Kidd and Watson's (1992) data suggest that a difference in lengths has a smaller effect if the lengths that differ are large but the task used imposes an upper bound on performance since one cannot discriminate between tone patterns with an accuracy above 100%. It would therefore be interesting to examine the issue using the Greenberg and Jenkins (1964) version of the SSJ task where numbers outputted by subjects can be arbitrarily large.

We still don't know much about the interaction of determinants of similarity with match and mismatch types. Neither is there much data regarding the possible explanations for the observed differences between languages and segments (an important exception here is Steriade 2004). In particular, can the emergence of sublexical units be explained by their acoustic properties (Steriade 2004) and/or the statistical regularities

available in the lexicon (Kessler and Treiman 1997, Yoon and Derwing 2001)? That is, what are the relevant input-based factors, and can they predict the emergence of all and only sublexical units in the languages of the world?

Finally, is sound similarity assessed in the same way across tasks? So far, there is remarkable agreement between sound-based tasks, including sound similarity judgments, AXB tasks, string manipulation, concept formation, unit counting, and same-different judgments. However, there are important differences in how position within the word impacts similarity in these tasks versus orthographic priming. It is important to determine how similarity is assessed in tasks measuring morphological productivity. Albright and Hayes (2003) and Experiment III in this paper are first steps in this direction.

## **CHAPTER 4. EXPERIMENT I: DEFINING ATTRIBUTES OF THE RULE-BASED DEFAULT ARE DISSOCIABLE<sup>29</sup>**

### ***4.1. Introduction***

The Dual Mechanism Model, or DMM, proposes that “regular” inflectional morphemes are attached by a categorical default rule understood as an operation over variables while “irregular” inflected words are formed by analogy to irregular forms in the lexicon (Pinker 1999, Pinker and Prince 1988, 1994, Berent et al. 1999, Clahsen 1999).

The two defining characteristics of a regular pattern applied via a default rule are 1) the regular pattern is the most frequently used pattern with nonce stems that are not similar to any existing stems, i.e. it is the pattern that is applied by default when formation by analogy fails; and 2) the regular pattern is applied when certain necessary and sufficient conditions defining the variables to which the rule applies have been satisfied, e.g. the stem is a verb, thus the regular pattern should be applied as readily to nonce stems that are similar to no existing stems as to stems similar only to existing stems taking the regular pattern, i.e. the regular pattern does not exhibit a similarity effect being insensitive to the contents of the nonce stimulus’ neighborhood.

The most serious challenge to DMM so far has come from Albright and Hayes (2003) who found that even the regular English past tense exhibits a similarity effect. Thus when native English speakers are asked to rate regular and irregular past tense forms of nonce verbs on how natural they sound as the past tense of the verb, they rate regular forms higher if the verb is similar to many regular verbs and no irregular ones than if it is

---

<sup>29</sup> This chapter has been presented at FASL 13, 4<sup>th</sup> International Conference on the Mental Lexicon (Kapatsinski 2004b), HDLS 6 (Kapatsinski 2004d), and BLS 31 (Kapatsinski 2005b). It appeared in the FASL 13 proceedings (Kapatsinski 2005).

similar to neither regulars nor irregulars. These results suggest that no pattern is free from similarity effects.

However, the DMM can account for these effects. Pinker and Prince (1994) stated that high frequency inflected forms are stored in the lexicon, whether they are regular or irregular. There is no theoretical mechanism that would prevent speakers from forming analogies based on these stored regulars. The analogical mechanism would then apply to all patterns but the rule would be a mechanism that is only used to attach regular patterns. Under this account, the default pattern does not have to be free of similarity effects but must simply be less susceptible to similarity effects than its competitor patterns. A stronger case against the DMM would be made if inflectional domains in which the pattern that is least susceptible to similarity effects is not the default pattern could be found. In this chapter, we show that Russian verbal stem formation is such a domain.

## ***4.2. Methods***

### **4.2.1. The task and the participants**

A written questionnaire was given to thirty-nine native Russian undergraduate students, studying at the University of the Russian Academy of Education in Nizhny Novgorod, Russia. The questionnaire contained 150 monosyllabic nonce consonant-final nouns in pseudorandomized order spelled in Cyrillic. The nouns were of six types: 1) similar to existing nouns taking –a- as the stem extension, 2) similar to existing nouns taking –i-, 3) similar to existing nouns taking –nu- 4) similar to existing nouns taking –{o;e}va-, 5) similar to existing nouns taking –e-, and 6) similar to no existing nouns. For instance, xɾʲuz → xɾʲuzatʲ, xɾʲuzitʲ, xɾʲuznutʲ, xɾʲuzovatʲ, xɾʲuzjetʲ.

The subjects were asked to make a verb out of the noun by filling in the missing part of the verb. The blank spaces in the middle of verbs were long to encourage the application of the longer stem extensions. The subjects were not given the final consonant of the verb root in order to encourage consonant changes and to ensure that maintenance of the consonant in environments where it could change was not due just to the subjects' reluctance to cross out letters printed in by the experimenter, e.g. (*pljuk*) *plju\_\_\_t'*, (*xrjuz*) *xrjuz\_\_\_t'*, (*lab*) *lab\_\_\_t'* (velars always change into palatals before stem extensions beginning with front vowels in the lexicon). Participants were shown examples orally prior to performing the task. The participants were not told what the choices were or what the purpose of the experiment was until after completing the experiment. Participants were under time pressure to complete the questionnaire and were told to write down the first thing that came to mind.

#### **4.2.2. Stimuli**

##### **4.2.2.1. Measuring phonological similarity**

Below is a summary of the similarity conventions adopted:

- limit = 4
- feature change = insertion or deletion of any number of C's on one edge = 1; if no C's remain, = 5
- insertion or deletion of a V or C word-internally = 5
- change of a phoneme keeping the phone = change of a phone keeping the phoneme = 0.5
- identical phonological shape = 0.25 (not 0 in order for division to be possible)

- category attractiveness =  $\sum_w \frac{1}{S}$  where  $w$  is a word obeying a certain pattern, while  $S$  is the similarity score ( $0.25 \leq S \leq 4$ )<sup>30</sup>
- $i$  was classified as a distinct phoneme rather than an allophone of  $i$

#### 4.2.2.2. Controlling for priming

It is always a danger when working with a large set of stimuli that the exposure to the preceding stimuli will affect the reactions to the following stimuli. In order to decrease the likelihood of such priming effects we pseudorandomized the order of presentation of the stimuli using the random number generator from random.org<sup>31</sup> and ensured that 1) no more than two stimuli close to words taking the same extension occurred consecutively, and 2) no more than two stimuli ending in the same consonant occurred consecutively.

Five stimuli from the very beginning of the experiment were repeated at the very end. The large number of intervening stimuli makes identity priming between the stimuli less likely. Thus, if there are significant differences between the two sets of stimuli, we may assume that they are to a large degree due to exposure to the intervening stimuli. By comparing the two sets of identical stimuli we can see if priming effects occurred and in what direction they were likely to bias the results. No statistically significant differences (at the .05 level according to the chi-square test) between first and second presentations of the same stimulus were found, either in the aggregate or in specific stimuli. Thus we can conclude that between-stimulus priming did not have a strong influence on the results.

---

<sup>30</sup> Only for three words was there a need to calculate category attractiveness, since for most words similarity and the number of similar words in that category were in agreement.

<sup>31</sup> This procedure was used in all experiments below.

#### 4.2.2.3. Root-final consonant

Gor and Chernigovskaya (2001) conducted an elicited production experiment where subjects were asked to produce the past tense form of a Russian verb given the present first person singular. They observed that the productivity of a stem extension was influenced by whether or not it required changes in the root. Shvedova et al. (1980) state that the final consonant of the nominal root changes before stem extensions that begin with a front vowel if it is a velar or a [ts], as shown in (1)-(3)<sup>32</sup>. This rule or schema is morphologized, that is, it applies to stem extensions but not case endings (cf. [tr<sup>j</sup>e 'voga] ‘worry’; [tr<sup>j</sup>e 'vog<sup>j</sup>i] ‘worries; of worry’; but [tr<sup>j</sup>e 'voʒut<sup>j</sup>] ‘to worry’<sup>33</sup>). However, the schema has no lexical restrictions.

(1) {[k]; [ts]} > [tʃ]

(2) [x] > [ʃ]

(3) [g] > [ʒ]

Thus, we should separate roots that end in a velar or [ts] from other roots, since we would expect stem extensions beginning with a front vowel to apply to these roots less often than to other roots.

The stimuli were balanced for palatalization, manner of articulation and sonority of the final consonant as well as the preceding vowel.

Table 4.1 shows the stimuli used in the experiment

---

<sup>32</sup> this rule is obligatory before stem extensions beginning with front vowel but is also often applied before back vowels, e.g. [be 'zdel<sup>j</sup>nik]/ [be 'zdel<sup>j</sup>nitʃat<sup>j</sup>]

<sup>33</sup> [i] changes to [ɨ] after [ʒ]

**Table 4.1. Stimuli**

Similarity Place	I	A	E	VA	NU	None
Velar and /ts/	l̥ets uk grak strok tʃerts kruk	mik lug skuk ʃix vorts brits	galk gr̥iek br̥iesk xraps svisk vl̥ʒk	bzix kuk lik garts strak vzbunk	bl̥iuk stvx pl̥iuk sog sdok bl̥iak	trapsk sprajk l̥ox ʒrak xvuts enk
Labial	sob jup̥i krip bamb gram straf	nuf̥ borv̥i s̥iejf l̥ev bl̥iam xlob	sum bim ʒum um gusp r̥ʒab	birb tr̥ieb zim l̥ub tramp ʃvarp	ʃep kr̥iep lib vsxl̥iep̥i vzdrim vspl̥ief	drajv̥i brisp ʃlurb xmap tl̥iup klorf̥
Other	xutʃ stul vol bor voj pl̥iuf̥i uʒ tʃan graz tvar	xritʃ xr̥iuz liss p̥Jes tif stril̥i tʃiz striʒ xliʃi skar	torst b̥iert bal ʒol grazd sr̥iet p̥ietr ogr̥i gl̥iat̥i ʒert̥i	notʃ ʃur dzot bomʒ vratʃ lied b̥iunt stort fl̥iert d̥iekt	xr̥ias spl̥iuz pris̥i m̥iors sbalt tr̥iaʃ vzdr̥ied fsplis br̥iz smort	ʃnit tejbl ʒnitʃ nutr̥i ajs ukl duvl̥i klift eks zr̥ieps

### 4.3. Results

#### 4.3.1. Palatalization of the root-final consonant

Palatalized consonants favor stem extensions beginning with front vowels and non-palatalized consonants disfavor them (Table 4.2). Table 4.3 shows that this result cannot be accounted for by claiming that palatalized root-final consonants occurred in words similar to existing words taking front vowel-initial stem extensions more than did non-palatalized root-final consonants, which obviates the need for controlling for similarity as we have done when testing the effect of sonority. Thus, palatalization has a

significant and independent effect.

**Table 4.2. Palatalization**

- sonorant		-cons; -back	other	Total
Labial p<0.000001	+pal	165 73.7%	59 26.3%	224
	-pal	488 57.2%	365 42.8%	853
	Average	326.5 60.1%	212 39.9%	538.5
Velar and <i>ts</i>	-pal	492 37.8%	811 62.2%	1303
Other p<0.0001	+pal	427 66.7%	213 33.3%	640
	-pal	650 56.6%	499 43.4%	1149
	Average	538.5 60.2%	356 39.8%	894.5
Average (excluding velars) p<0.00001	+pal	592 68.5%	272 31.5%	864
	-pal	1138 56.8%	864 43.2%	2002
	Average	865 60.4%	568 39.6%	1433

**Table 4.3: Palatalized Root-final consonants by stimulus similarity**

Close to—> Have	-cons; -back	+back or +cons	Neither
+palatal	6 12.5%	8 18.2%	5 22.7%
-palatal	42 87.5%	36 81.8%	17 78.3%
Total	48	44	22

Results reported in the next section are independent of the effect of palatalization, since only stimuli with non-palatalized final consonants were analyzed.

#### 4.3.2. Similarity effects and productivity

Table 4.4. shows that -a- is the pattern that is used more frequently than any other pattern with velar-final roots that are not similar to any existing roots (chi-square=5.37,

p<0.025). This is in agreement with the observation that stem extensions beginning with front vowels do not come after velar-final stems in the lexicon. Furthermore, -a- is the only pattern that does not exhibit a numerical similarity effect in that it is applied to stimuli that are different from all existing roots as frequently as it applies to stimuli that are similar to roots preceding -a-. Since the magnitude of -a-'s similarity effect is significantly different from the magnitude of its nearest competitor's, -i-'s similarity effect (chi.sq.=24, p<0.001), -a- is the pattern that exhibits the weakest similarity effect.

We may note, however, that the stem extension -i- was applied after velar-final roots in 33% of the responses (chi.sq.=21.2, p<0.001), which indicates that even schemas that are true for every lexical item in the lexicon that meets their input specifications are not necessarily fully productive (cf. Zimmer, 1969, for Turkish). This may be because the schema in question is morphologized.<sup>34</sup>

**Table 4.4. Affix productivity after velars. X stands for ‘the class of roots that take the stem extension shown on the right’.**<sup>35</sup>

X	I	A	E	{o;e}VA	NU
Similar to X	44% 98/222	48% 106/222	10% 22/222	23% 52/222	13% 29/222
Similar to none	33% 74/222	47% 105/222	4% 8/222	9% 19/222	5% 12/222
Significance	no	no	p<0.05	p<0.001	p<0.01
Similarity effect	11%	1%	6%	14%	8%

Table 4.5 shows that -a- is the pattern that is used more frequently than any other pattern with labial-final roots that are not similar to any existing roots (although the difference between -a- and -i- is not statistically significant). -a- again appears to behave as the default pattern in terms of similarity effects (the differences between -a-'s reverse

<sup>34</sup> It is not lexicalized because it occurs with all roots.

<sup>35</sup> Percentages indicate the number of responses choosing the stem extension shown in the top row of the column. Note that the percentages in this table need not add up to 100 either horizontally in the top row or vertically because the percentages come from different groups of stimuli (see appendix for complete results).

similarity effect and no similarity effect displayed by its nearest competitors is significant,  $\chi^2=26$ ,  $p<0.001$ ).

**Table 4.5. Affix productivity after labials.**

X	I	A	E	{o;e}VA	NU
Similar to X	62% 115/185	26% 39/148	14% 25/185	19% 35/184	7% 11/148
Similar to none	36% 53/148	44% 65/148	8% 12/148	5% 8/148	4% 6/148
Significance	$p<0.005$	$p<0.025$	no	$p<0.001$	No
Similarity effect	26%	-18%	6%	14%	3%

Table 4.6 shows that while -a- is the only affix that exhibits no similarity effect after coronal and palatal consonants: the difference between -nu- and -a-'s similarity effects is significant ( $\chi^2=5$ ,  $p<0.05$ ) the difference between the similarity effects shown by -a- and -i- is significant at  $p<0.001$  ( $\chi^2=33$ ), -i- is the affix used most after roots that are not similar to any existing roots ( $\chi^2=5.14$ ,  $p<0.025$ ). Thus the two characteristics of a default proposed within the DMM are dissociable.

**Table 4.6. Affix productivity after coronals and palatals**

X	I	A	E	{o;e}VA	NU
Similar to X	63% 139/222	32% 72/222	20% 75/370	18% 54/296	9% 27/296
Similar to none	42% 109/259	30% 78/259	8% 21/259	10% 27/259	4% 9/259
Significance	$p<0.005$	No	$p<0.001$	$p<0.025$	$p<0.001$
Similarity effect	21%	2%	12%	8%	5%

#### **4.4. Discussion**

Contra the DMM (Clahsen 1999, Pinker 1999), the most productive affix is not necessarily the one that is least affected by the similarity of the nonce stimulus to existing stems. In the following chapters we explore the factors driving the size of the similarity effect and productivity.

#### 4.5. Appendix: Results in detail

**Table 4.7. Velar-final roots<sup>36</sup>**

Similar to → Takes ↓	I	A	E	VA	NU	None
I	98 44.1%	57 25.7%	60 27.0%	63 28.4%	47 21.2%	74 33.3%
A	85 38.3%	106 47.7%	85 38.3%	83 37.4%	113 50.9%	105 47.3%
E	7 3.2%	12 5.4%	22 9.9%	12 5.4%	6 2.7%	8 3.6%
VA	17 7.7%	28 12.6%	23 10.4%	52 23.4%	14 6.3%	19 8.6%
NU	12 5.4%	12 5.4%	16 7.2%	9 4.1%	29 13.1%	12 5.4%
Other	12 5.4%	7 3.2%	17 7.7%	3 1.4%	13 5.9%	4 1.8%
Total	222	222	222	222	222	222

**Table 4.8. Labial-final roots**

Similar to	I	A	E	VA	NU	None
I	115 62.2%	66 44.6%	85 45.9%	67 36.4%	63 42.6%	53 35.8%
A	36 19.5%	39 26.4%	49 26.5%	52 28.2%	48 32.4%	65 43.9%
E	16 8.6%	12 8.1%	25 13.5%	21 11.4%	9 6.1%	12 8.1%
VA	9 4.9%	15 10.1%	7 3.8%	35 19.0%	12 8.1%	8 5.4%
NU	8 4.3%	9 6.1%	12 6.5%	7 3.8%	11 7.4%	6 4.1%
Other	1 0.5%	7 4.7%	7 3.8%	3 1.6%	5 3.4%	4 2.7%
Total	185	148	185	185	148	148

<sup>36</sup> Note that in Kapatsinski (2005) one velar-final stimulus was misclassified as similar to –a- when it was similar to –i-, one stimulus similar to –nu was misclassified as similar to –Vva and one stimulus similar to –e was omitted because of its high rate of ‘other’ affixation. We now reclassified the misclassified stimuli and included the stimulus similar to –e. Also some typos in the tables were fixed through comparison with the raw data. Cases in which subjects made no response were formerly excluded and now included as ‘other’. These changes did not lead to changes in the pattern of results.

**Table 4.9. Coronal- and palatal-final roots**

Similar to	I	A	E	VA	NU	None
I	139 62.6%	109 49.1%	167 45.1%	136 45.9%	110 37.2%	109 42.1%
A	41 18.5%	72 32.4%	80 21.6%	53 17.9%	106 35.8%	78 30.1%
E	13 5.9%	17 7.7%	75 20.3%	32 10.8%	33 11.1%	21 8.1%
VA	11 5.0%	10 4.5%	31 8.4%	54 18.2%	16 5.4%	27 10.4%
NU	6 2.7%	10 4.5%	14 3.8%	7 2.4%	27 9.1%	9 3.5%
Other	12 5.4%	4 1.8%	3 0.8%	14 4.7%	4 1.4%	15 5.8%
Total	222	222	370	296	296	259

## CHAPTER 5. CORPUS STUDY I: EVENNESS OF DISTRIBUTION AND SENSITIVITY TO SIMILARITY<sup>37</sup>

### *5.1. Introduction*

This chapter tries to answer the question of why -a- is less sensitive to the contents of the nonce stimulus' phonological neighborhood than -i- after roots with any place of articulation of the root-final consonant. Hare et al. (1995) have shown that in connectionist networks low sensitivity to phonological neighbors is predicted to occur with affixes that occur in words that are distributed relatively evenly throughout the lexicon. Thus, if a morpheme attaches to words that are very different from each other, it will not be strongly associated with any particular phonological and semantic characteristics of the words and would therefore exhibit little sensitivity to what a nonce stimulus is similar to. Boudelaa and Gaskell (2002) have demonstrated that nouns containing the proposed minority default plural in Arabic are distributed more evenly in phonological space than words containing other plurals, which accounts for smaller similarity effects exhibited by the 'default' morpheme. In this chapter, we show that the same is true in Russian: -a-containing words are more phonologically diverse and -a- exhibits the smallest similarity effect.

### *5.2. Methods*

In order to determine how evenly -i- and -a- are distributed, I took all verbs listed in the reverse dictionary of Russian (Sheveleva et al. 1974)<sup>38</sup> that contained either -i- or -a- as

---

<sup>37</sup> This chapter was presented at HDLS VI (Kapatsinski 2004d) and BLS 31 (Kapatsinski 2005b).

<sup>38</sup> The same set of words is contained in Zaliznjak (1977), used in experiment I, and Sheveleva et al. (1974), used here.

the stem extension. Those verbs in which -a- followed a palatalized consonant were included in the counts of verbs containing –a. A total of 4512 verbs were analyzed.

For root-final, penultimate and antepenultimate positions, the number of words having a particular phoneme in the position was calculated. Separate calculation were performed for every place of articulation of the root-final consonant. For each place of articulation of the root-final consonant, the phonemes in a particular position within the root were arranged on a bar graph from the most common to the least common. The distribution closely approximated a logarithmic curve. Therefore, a logarithmic curve was used as a trendline. The slopes of the curves representing the distribution of phonemes in a particular position across the roots of words containing -a- vs. -i- were then compared. In the comparison, the steeper the line (the larger the absolute value of the slope), the less evenly the words containing an affix are distributed in phonological space.

### ***5.3. Results and discussion***

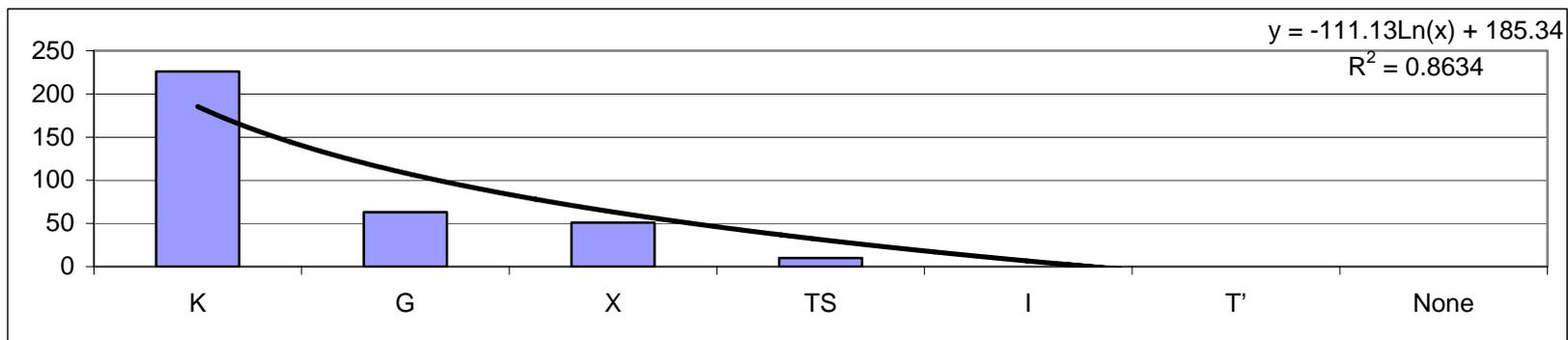
Table 5.1. shows the slopes for the lines shown in figures 5.1-5.15. As can be seen, for all root-internal positions, words containing -a- whose roots end in a labial, a coronal, or a palatal are distributed more evenly in phonological space than word containing -i-, leading us to expect a smaller similarity effect exhibited by -a- for these classes of roots. However, for roots ending in a velar, -a-containing words predominantly have /k/ as the root-final consonant. Nevertheless, we did not find a similarity effect.

**Table 5.1. Slopes of lines in figures 5.1-5.15, where the smaller the absolute value of the slope, the more evenly the words containing the stem extension are distributed in phonological space<sup>39</sup>**

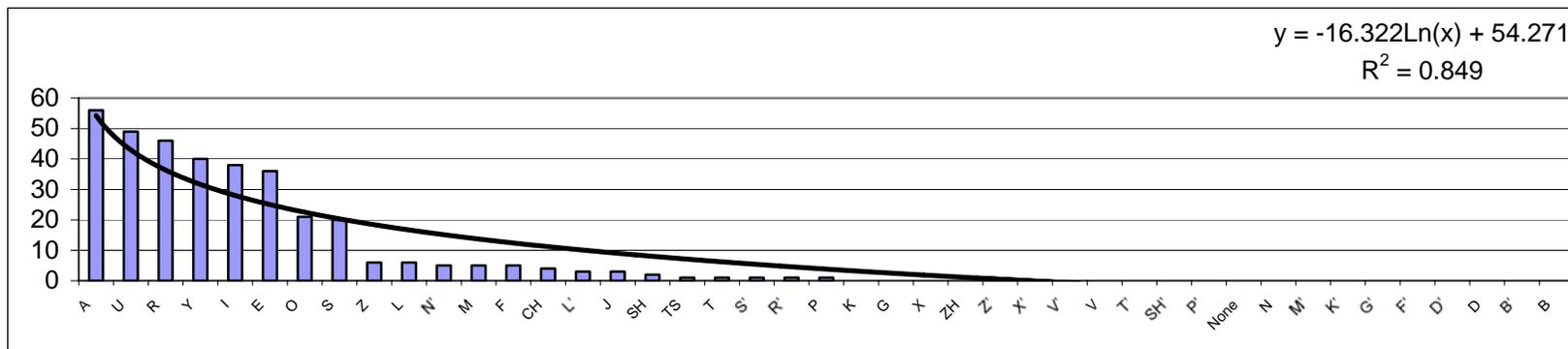
In verbs bearing → Phoneme location ↓	-a-	-i-
Final in velar-final roots e.g. <b>rug</b> , <b>lak</b> , <b>sox</b>	111.1	0
Penultimate in velar-final roots e.g. <b>rug</b> , <b>lak</b> , <b>sox</b>	16.3	
Antepenultimate in labial-final roots e.g. <b>rug</b> , <b>lak</b> , <b>sox</b>	8.6	
Final in labial-final roots e.g. <b>lom</b> , <b>kop</b> , <b>rub</b> , <b>sov</b> , <b>traf</b>	20.1	39.9
Penultimate in labial-final roots e.g. <b>lom</b> , <b>kop</b> , <b>rub</b> , <b>sov</b> , <b>traf</b>	5.2	13.2
Antepenultimate in labial-final roots e.g. <b>lom</b> , <b>kop</b> , <b>rub</b> , <b>sov</b> , <b>traf</b>	3.5	8.1
Final in coronal and palatal-final roots e.g. <b>xod</b> , <b>mot</b> , <b>kos</b> , <b>voz</b> , <b>roʒ</b> , <b>vʲeʃ</b> , <b>vʲeʃʲ</b>	41.5	84.0
Penultimate in coronal and palatal-final roots e.g. <b>xod</b> , <b>mot</b> , <b>kos</b> , <b>voz</b> , <b>roʒ</b> , <b>vʲeʃ</b> , <b>vʲeʃʲ</b>	38.1	62.0
Antepenultimate in coronal and palatal-final roots e.g. <b>xod</b> , <b>mot</b> , <b>kos</b> , <b>voz</b> , <b>roʒ</b> , <b>vʲeʃ</b> , <b>vʲeʃʲ</b>	17.4	23.8

<sup>39</sup> Palatalized consonants were counted as different from non-palatalized ones. Non-palatalizing and palatalizing vowels were counted as the same when acoustically identical, i.e. except the /i/-/i/ contrast.

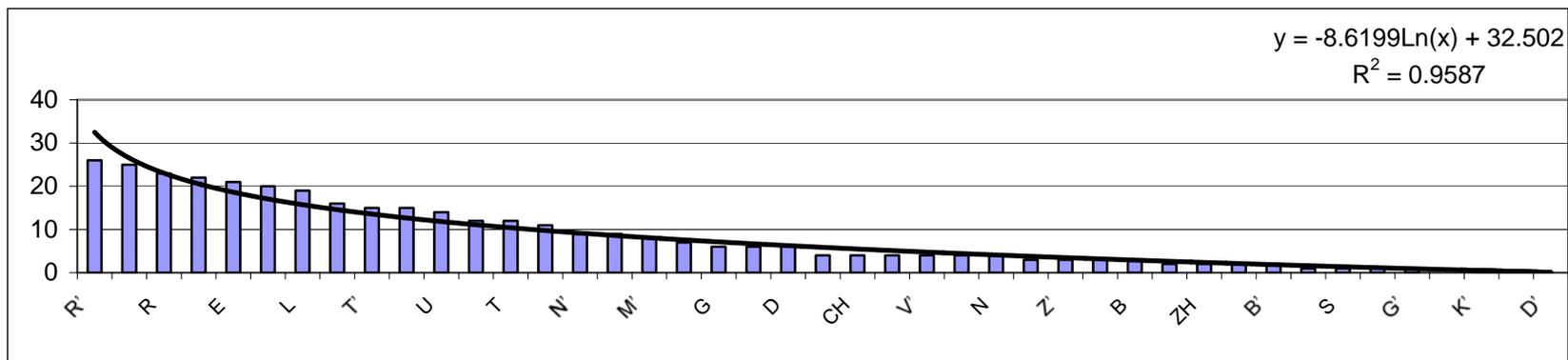
**Figure 5.1. Type frequencies with which different velar consonants occur in the root-final position in verbs containing -a-**



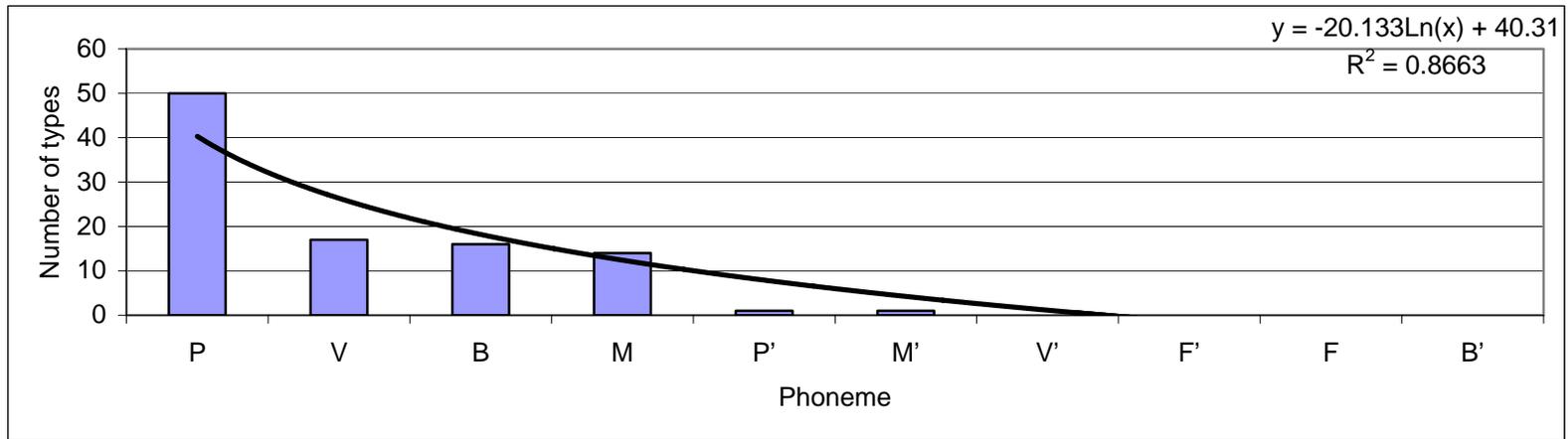
**Figure 5.2. Type frequencies with which different sounds occur in the penultimate position in verbs containing -a- that have a velar root-final consonant.**



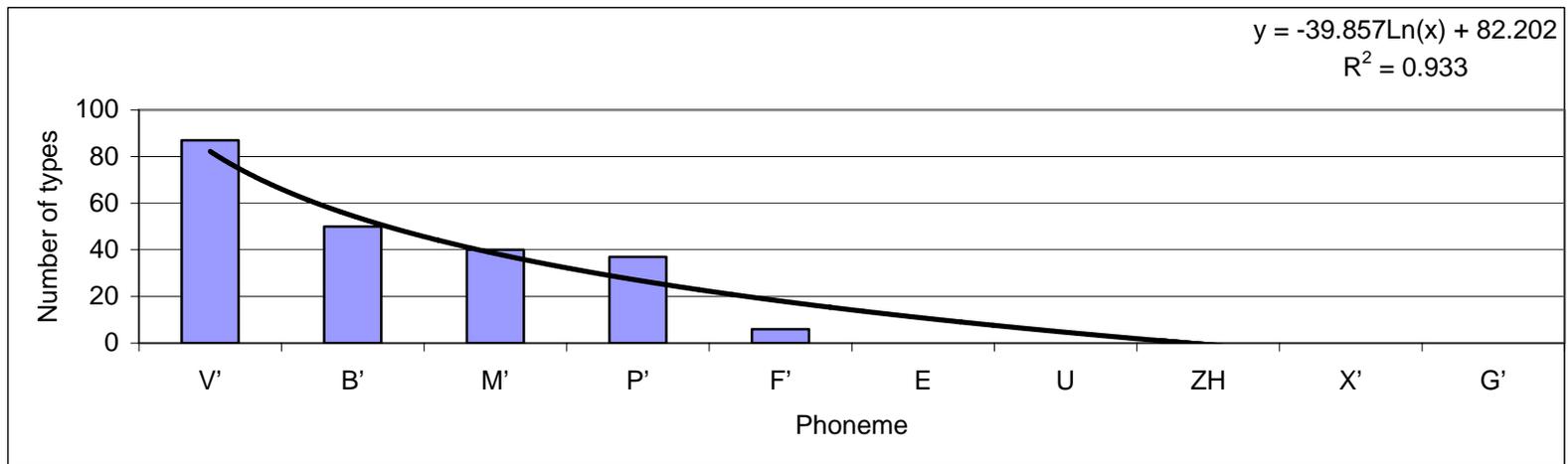
**Figure 5.3. Type frequencies with which different sounds occur in the antepenultimate position in verbs containing -a- that have a velar root-final consonant.**



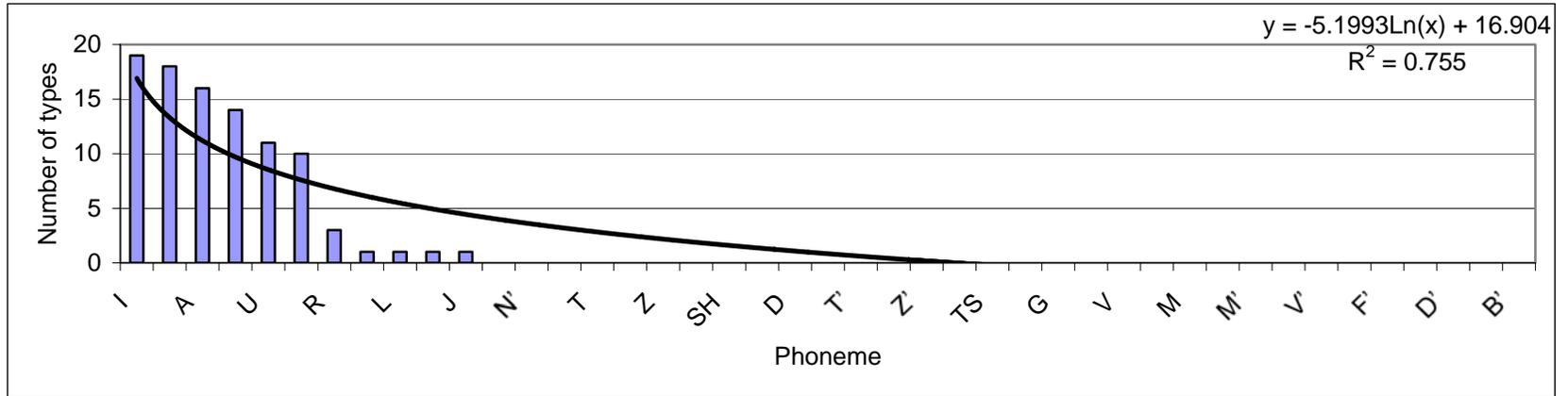
**Figure 5.4. Type frequencies with which different labial consonants occur in the root-final position in verbs containing - a-.**



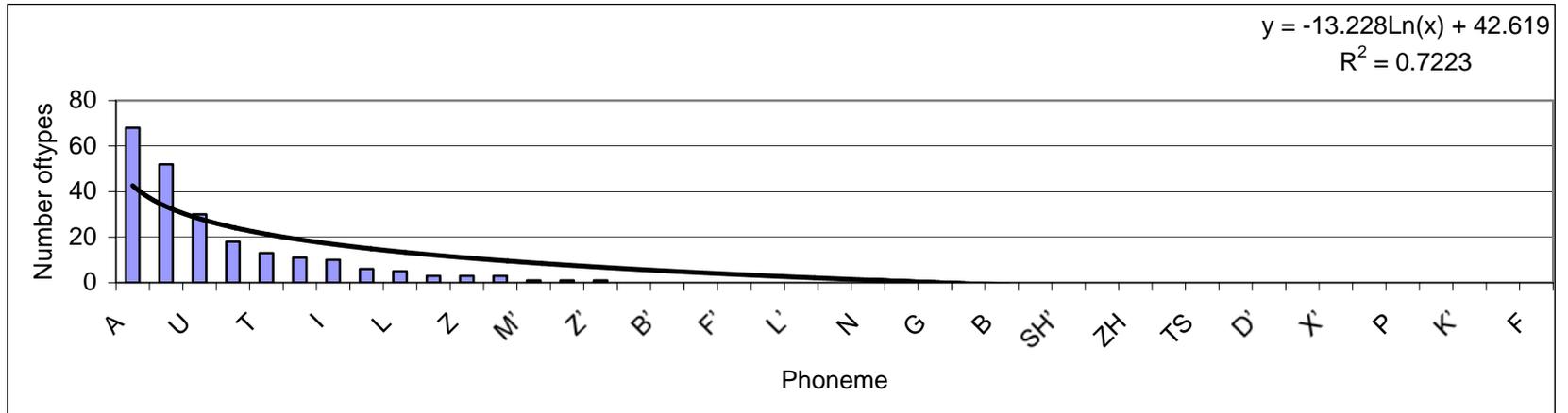
**Figure 5.5. Type frequencies with which different labial consonants occur in the root-final position in verbs containing - i-.**



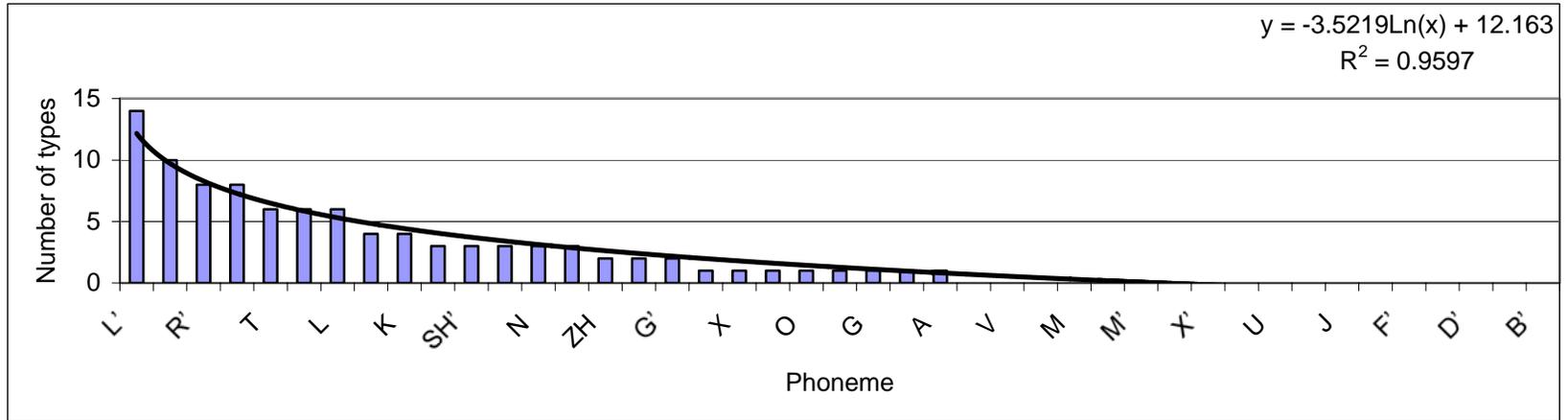
**Figure 5.6. Type frequencies with which different phonemes occur in the penultimate position in verbs containing -a- that have a labial root-final consonant.**



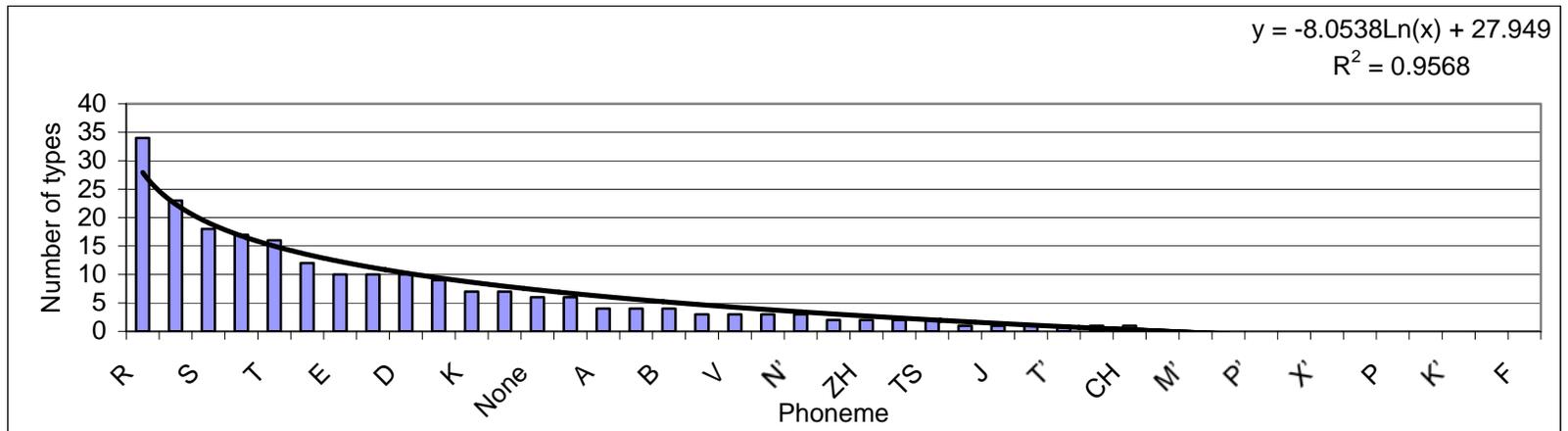
**Figure 5.7. Type frequencies with which different phonemes occur in the penultimate position in verbs containing -i- that have a labial root-final consonant.**



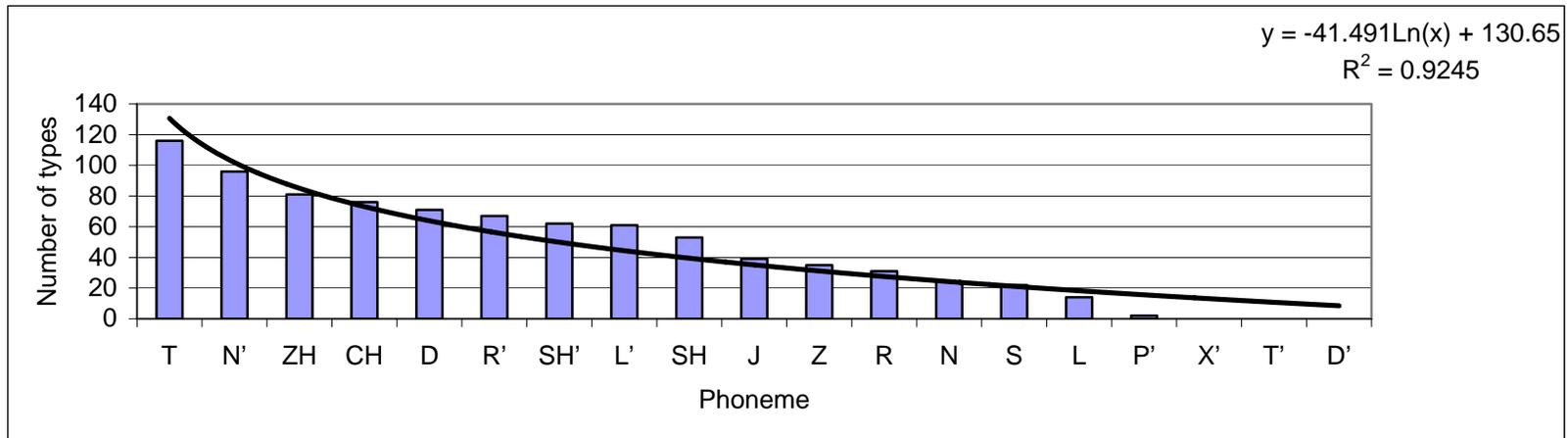
**Figure 5.8. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -a- that have a labial root-final consonant.**



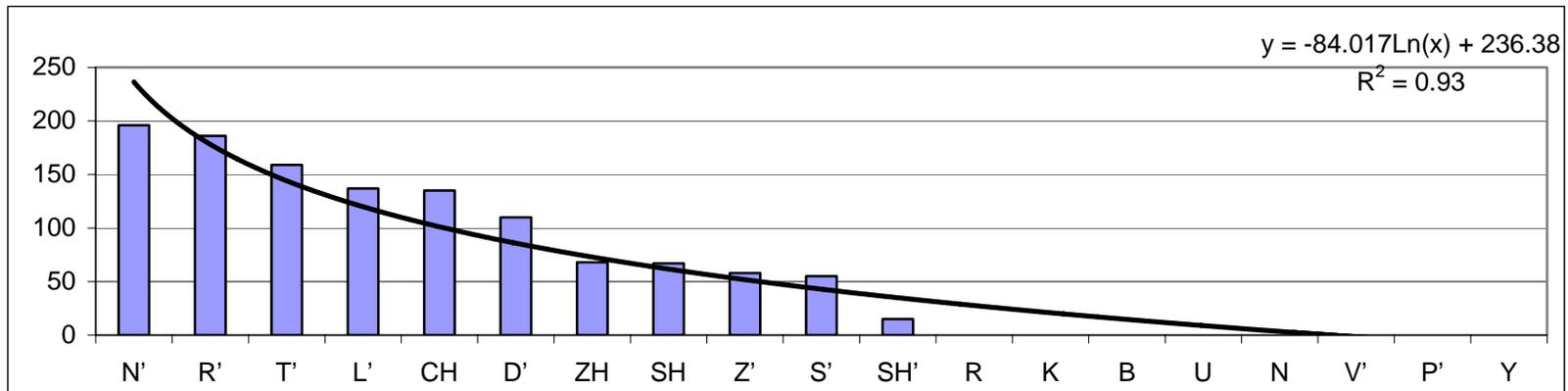
**Figure 5.9. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -i- that have a labial root-final consonant.**



**Figure 5.10. Type frequencies with which different coronal and palatal consonants occur in the root-final position in verbs containing -a-.**

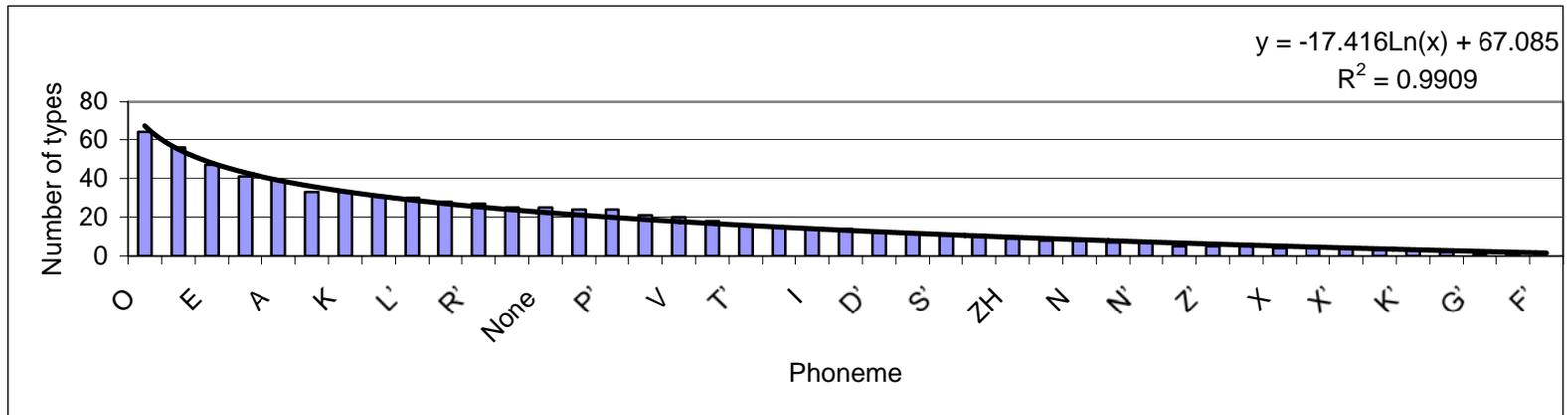


**Figure 5.11. Type frequencies with which different coronal and palatal consonants occur in the root-final position in verbs containing -i-.**

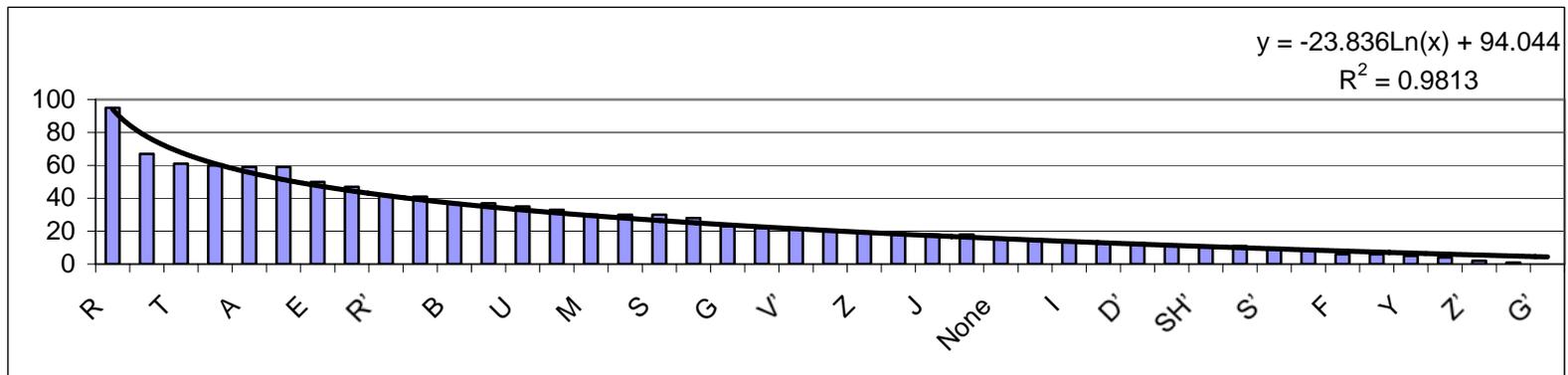




**Figure 5.14. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -a- that have a coronal or palatal root-final consonant.**



**Figure 5.15. Type frequencies with which different phonemes occur in the antepenultimate position in verbs containing -i- that have a coronal or palatal root-final consonant.**



Close examination of the stimuli, however, reveals that more stimuli that were similar to no existing words ended in /k/ than stimuli that were similar to word taking -a-. This explains why no similarity effect was observed for -a- and suggests that a match in the final consonant is weighted much more heavily than matches elsewhere, as suggested by studies of sound similarity (Derwing and Nearey 1986, Bendrien 1992, Kapatsinski 2004e, chapter 3 of this thesis) and/or there is a general schema in which /k/ is associated with a following -a-<sup>40</sup> or /a/<sup>41</sup> that impacts morpheme choice above and beyond the influence of neighbor-based analogy through co-occurrence-based priming. The next chapter reports on an experiment designed to examine these issues.

---

<sup>40</sup> The morpheme

<sup>41</sup> The phoneme

## **CHAPTER 6. EXPERIMENT II: ALL THE POPULAR SUFFIXES ARE SENSITIVE**

### ***6.1. Introduction***

In experiment I, we have seen that -a- seems to be less sensitive to the contents of the nonce stimulus's neighborhood. However, one factor that was not controlled in the experiment is the identity of the root-final consonants in stimuli similar to -a-bearing verbs and stimuli similar to -i-bearing roots. We have seen that the root-final consonant influences the relative productivity of the following suffixes in that palatalized root-final consonants favor stem extensions beginning with front vowels. In this experiment, we used minimal pairs of stimuli similar to no existing words that differed only in the root-final consonant to investigate the influence of the root-final consonant further. In addition, stimuli similar to -i-bearing verbs, those similar to -a-bearing verbs, and those similar to no existing verbs that are being compared in this experiment are matched on the identity of the root-final consonant. Unlike in experiment I, stimuli in this experiment were presented auditorily. Finally, instead of asking the subjects to come up with a verb given a noun, they were presented with noun-verb pairings and asked to rate how likely people are to form the verb from the noun on a 10-point scale. Albright and Hayes (2003) have previously found that for the English past tense ratings obtained in this task are highly correlated with the probability of producing the past tense in an elicited production task. Both -i- and -a- turned out to be sensitive to the contents of the nonce stimulus's neighborhood and, as long as the identity of the root-final consonant is controlled, sensitive to the same degree.

## **6.2. Methods**

Eighteen native Russian speakers were presented auditorily with a nonsense noun followed by a verb formed from it via either -i-, -a-, -ova-, or -eva-. They were asked to rate the pairs on a 10 point scale ranging from 1="nobody would form this verb from this noun" to 10="everyone would form this exact verb from this noun". Subjects were asked to imagine that the nouns are recent borrowings and they need a verb to describe an event involving the noun.

Minimal pairs of nouns similar to no existing nouns were created by varying the place of articulation of the final consonant. Thus, there were minimal pairs exemplifying contrasts between /b/ and /d/, between /p/ and /t/, between /v/, /z/, and /ʒ/, and between /m/ and /n/. Subjects' reactions to the members of the minimal pairs were compared to determine the influence of the root-final consonant independently of the influence of the neighbors and the non-final segments. Velar consonants were not used so as not to confound the necessity to perform an extra operation with the identity of the consonant. All consonants were non-palatalized when occurring noun-finally.

Stimuli similar to roots taking -a- and -i- were created so that the set of stimuli similar to -a-bearing words, the set of stimuli similar to -i-bearing words and the comparison set of stimuli similar to no existing words contained the same proportions of final consonant types. Thus, the sets of stimuli similar to -i-bearing and -a-bearing words each contained two nouns ending in /b/, 1 ending in /m/, 1 ending in /t/, and 1 ending in /p/. The set of stimuli similar to no existing words was twice as large, containing 4 /b/-final and 2 /p/-final, /m/-final, and /t/-final nouns. Thus any differences found between the sets cannot be attributed to the influence of the root-final consonants.

All stimuli similar to -a-taking words had no -i-taking words within the radius and vice versa. This was done to ensure that the magnitude of the differences between -a-favoring and -i-favoring words vs. neutral words are not due to an experimental confound but are rather a genuine reflection of differences in -a- and -i-'s sensitivity to the contents of the nonce stimulus's neighborhood.

Stress was assigned to the first syllable to half of the stimuli similar to no existing verbs and to the second syllable for the other half. It was assigned in a way that maximized similarity for the stimuli similar to -a-bearing and -i-bearing verbs. Stimuli are presented in tables 6.1 and 6.2.

**Table 6.1. Stimuli used to test the size of similarity effects or -a- and -i-<sup>42</sup>**

Stimuli bearing	Stimuli similar to -a-bearing words	Stimuli similar to -i-bearing words	Stimuli similar to no words
-a-	xli'bat <sup>ɨ</sup> .5 xlej'bat <sup>ɨ</sup> 1 dri'mat <sup>ɨ</sup> .5 'xrjupat <sup>ɨ</sup> 1 ri'tat <sup>ɨ</sup> .5	'grajbat <sup>ɨ</sup> 1 'zlubat <sup>ɨ</sup> 1 gra'mat <sup>ɨ</sup> .5 gluj'pat <sup>ɨ</sup> 1 'tortat <sup>ɨ</sup> 1	'kv <sup>ɨ</sup> ubat <sup>ɨ</sup> 'ʒrjubat <sup>ɨ</sup> 'tʃembat <sup>ɨ</sup> rem'bat <sup>ɨ</sup> ksa'mat <sup>ɨ</sup> ʃver'mat <sup>ɨ</sup> ʃej'pat <sup>ɨ</sup> x <sup>ɨ</sup> em'pat <sup>ɨ</sup> 'zjɨtat <sup>ɨ</sup> 'zviltat <sup>ɨ</sup>
-i-	xli'bit <sup>ɨ</sup> xlej'bit <sup>ɨ</sup> dri'mit <sup>ɨ</sup> 'xrjupit <sup>ɨ</sup> ri'tit <sup>ɨ</sup>	'grajbit <sup>ɨ</sup> 'zlubit <sup>ɨ</sup> gra'mit <sup>ɨ</sup> gluj'pit <sup>ɨ</sup> 'tortit <sup>ɨ</sup>	'kv <sup>ɨ</sup> ubit <sup>ɨ</sup> 'ʒrjubit <sup>ɨ</sup> 'tʃembit <sup>ɨ</sup> rem'bit <sup>ɨ</sup> ksa'mit <sup>ɨ</sup> ʃver'mit <sup>ɨ</sup> ʃej'pit <sup>ɨ</sup> x <sup>ɨ</sup> em'pit <sup>ɨ</sup> 'zjɨtit <sup>ɨ</sup> 'zviltit <sup>ɨ</sup>

<sup>42</sup> The stimuli are presented in a phonological IPA transcription.

**Table 6.2. Stimuli used to test the influence of the root-final consonant**

Stem extensions tested	Root-final consonant		
-a-, -i-	v	z	ʒ
	gliv glerv fr <sup>h</sup> uv xejv xmev	gliz glerz fr <sup>h</sup> uz xejz xmez	gliʒ glerʒ fr <sup>h</sup> uʒ xejʒ xmeʒ
-a-, -i-, -ova-, -eva-	b	d	
	gl <sup>h</sup> uzb dvizb ʒr <sup>h</sup> ub zmib kv <sup>h</sup> ub remb tʃemb	gl <sup>h</sup> uzd dvizd ʒr <sup>h</sup> ud zmid kv <sup>h</sup> ud rend <sup>43</sup> tʃend	
	p	t	
	zvilp zijp m <sup>h</sup> orp stvurp x <sup>h</sup> emp ʃejp	zvilt zijt m <sup>h</sup> ort stvurt x <sup>h</sup> ent ʃejt	
	m	n	
	f <sup>h</sup> v <sup>h</sup> erm xv <sup>h</sup> ejm flim ksam ʒlum	f <sup>h</sup> v <sup>h</sup> ern xv <sup>h</sup> ejn flin ksan ʒlun	

Stimuli were presented in a pseudorandomized order using a random sequence from random.org. Locations in a second random sequence were used to assign positions to the first of two members of a minimal pair or two stimuli similar to words taking the same stem extension that were assigned by the first sequence to be within 10 stimuli of each other. In addition, it was ensured that the first 10 stimuli were similar to no existing

<sup>43</sup> Nasals undergo place assimilation before obstruents.

words and that they were followed by one stimulus similar to -a- and one similar to -i-. The order of the -a-favoring and -i-favoring stimuli was switched for half of the subjects. There was no significant difference between the resulting subject groups.

Each stimulus was presented once. Subjects were tested in groups. One female native Russian speaker pronounced all of the stimuli for all the subjects. The subjects were not presented with orthographic representations of the stimuli. The scales were drawn on response sheets.

The data were analyzed using a univariate ANalysis Of VAriance (ANOVA) in SPSS. Respondent identity was entered into the ANOVA as an independent variable. In addition, all stimuli were coded for stress location since half of the stimuli similar to no existing words were stressed on the root and half were stressed on the affix while stress location in stimuli similar to -a-bearing or -i-bearing words was chosen to maximize the number of same-category members.

### ***6.3. Results and discussion***

As Table 6.3. shows, both -a- and -i- show robust similarity effects. Moreover, the sizes of the effects seem to be relatively equal. If this is indeed the case, we are forced to conclude that the more even lexical distribution of -a- makes it harder to find similarity effects for -a- methodologically but does not make -a-'s application less reliant on the contents of the stimulus's neighborhood.

**Table 6.3. Acceptability of nonce verbs bearing -a- and -i- depending on the contents of their phonological neighborhoods**

The stimulus bears → Neighbors bear▼	A	I
A	6.27	5.16
I	4.73	6.40
No neighbors	4.70	5.05
Significance	p<0.0005	p<0.0005
Size of the similarity effect =A/No Neighbors	1.33	1.27

However, while the final consonant has been controlled in the dataset explored in table 6.3, some stimuli similar to -a- or -i- are 0.5 UNOS removed from the nearest neighbor and some are 1 UNOS removed. A central prediction of the FIRM is that the internal structure of the neighborhood is graded. That is, distance from the stimulus within the neighborhood influences the similarity of the neighbor to the stimulus, as opposed to all neighbors within the radius being equally similar to the stimulus. Stimuli similar to -a- contain 3 stimuli 0.5 UNOS removed from their -a-bearing neighbors, while stimuli similar to -i- contain only 1 stimulus 0.5 UNOS removed from its -i-bearing neighbor. Therefore, we may be overestimating -a-'s similarity effect in relation to -i-.

Table 6.4 presents the mean scores for stimuli similar to -i- and -a- and stimuli similar to no existing roots as a function of the distance between the stimulus and the nearest possible analogical model. At every distance from the nearest analogical model, -i- is the more productive of the two affixes. However, the difference is only significant when the nearest neighbor is 0.5 UNOS away ( $p=0.009$ ). These results suggest that -a- may be slightly less sensitive to similarity than -i-, as found in experiment I. However, only one of the stimuli similar to -i- has a neighbor within 0.5 UNOS. Thus the data sample is small and we should be cautious in making this inference. Whether -a- is

slightly less sensitive to similarity than -i- or not, it is clear that -i- is not less sensitive to similarity than -a- and is not less productive than -a-. Thus, the findings directly contradict the Dual Mechanism Model.

**Table 6.4. Distance to nearest neighbor and sensitivity to similarity**

Similar to verbs bearing	Distance to neighbor	-a-bearing stimuli	-i-bearing stimuli
-a-	0.5	6.89 dif. from 1: p=0.005	5.13
	1	5.65 dif. from none: p=0.005	5.19
-i-	0.5	5.28	8.22 dif. from 1: p<0.0005
	1	4.56	5.89 dif. from none: p=0.011
None		4.70	5.05
Similarity effect (0.5/None)		1.5	1.6
Similarity effect (1/None)		1.2	1.2

Because 3 stimuli labeled as similar to verbs bearing -a- are acoustically identical to real verbs bearing -a-, and 1 stimulus similar to -i-taking verbs is acoustically identical to an -i-taking verb (i.e. have a radius of .5 in table 6.1), one could hypothesize that the similarity effects observed are simply due to these verbs being perceived as real verbs bearing -i- and not due to analogy at all. If this were the case, there would be no difference between -a-bearing stimuli similar to -a-taking verbs for which the nearest neighbor is 1 UNOS away and -a-bearing stimuli similar to verbs taking -i-. Similarly, there should be no difference between -a-bearing stimuli similar to -a-taking verbs for which the nearest neighbor is 1 UNOS away and -a-bearing stimuli similar to verbs taking -i-. However, for both -a- and -i-bearing verbs, there is a significant difference between stimuli 1 UNOS removed from a model and those with no model within 4 UNOS (Table 6.4). Thus the results do indicate that similar verbs, and not just identical

ones, are taken into account.

Since -i- is not less productive than -a- and is not less sensitive to similarity, the DMM is forced to hypothesize that -i- and -a- must be less productive than some other stem extension, which must be less sensitive to similarity than both of them. The only plausible alternative is -ova-/-eva-, since the stem extension is reasonably frequent and is often applied to borrowings. Table 6.5 shows the relative naturalness ratings of stimuli derived from the same set of nouns similar to no existing words with -i-, -a-, -ova-, and -eva-. Contrary to the Dual Mechanism Model's prediction, -i- and -a- are the stem extensions that derive the most natural-sounding verbs. The lower acceptability of -ova-/-eva-bearing words is due to their greater length (cf. Coleman and Pierrehumbert 1997, Frisch et al. 2000 for nonsense words in English). The acceptability of -a- and -i-bearing stimuli may also be artificially inflated by presenting subjects with stimuli similar to -i- and -a-bearing verbs but not to -ova- and -eva-bearing verbs. However, the same result is obtained with a standard elicited production task (Experiment I). The difference between -a- and -ova-/-eva- is highly statistically significant ( $p < 0.0005$ ) when assessed in an ANOVA that also includes respondent identity, stress location, and final consonant identity. The difference between -i- and -a- is also significant when tested in the same way ( $p = 0.008$ ). Nonetheless, the similarity effects displayed by -i- and -a- are relatively equal. The difference between -ova- and -eva- is not significant ( $p = 0.564$ ).

**Table 6.5. Naturalness ratings of (coronal- and palatal-final) verbs derived from nouns similar to no existing words with -a-, -i-, -ova-, or -eva-.**

Stem extension	-i-	-a-	-ova-	-eva-
Mean rating	5.28	4.97	4.32	4.30

Tables 6.2 and 6.3 show that verbs bearing -a- are just as acceptable when they

are similar to -i-bearing verbs as they are when they are similar to no existing words. Similarly, high similarity to -a-bearing verbs and lack of similarity to any existing verbs have the same effect on -i-bearing verbs. This suggests that the differences in judgments are not due to lower phonotactic probabilities of words that are similar to no existing words. In addition, it appears that the judgment is based solely on the absolute number of words in the stimulus's neighborhood that bear the stem extension the stimulus bears, rather than on the number of words in the stimulus's neighborhood with the same stem extension relative to the total number of words in the neighborhood: stimuli similar to words bearing other stem extensions and stimuli similar to no words are similar to no words bearing the stem extension of the stimulus but stimuli that are similar to the 'wrong' stem extension might be expected to reduce the acceptability of the stimulus if stem extensions are able to compete. For instance, if the judgments are supposed to be based on the relative activation levels of the stem extension borne by the stimulus relative to the resting activation levels of other stem extensions that the stimulus could bear but didn't. Despite the instructions, which explicitly mentioned estimating the likelihood of producing this *exact* verb given the noun, subjects seem to have based their judgments solely on the activation level of the stem extension borne by the stimulus and not on the activation levels of other extensions.

One might hypothesize that the finding is due to the stimuli containing the stem extension on which the judgment is made. If this were the case, the finding would be irrelevant for morphological productivity or production tasks. In such tasks, the competition is enforced because a subject must choose only one alternative. Therefore, regardless of whether the numerical relationship between the competitors in the

stimulus's neighborhood matters, higher attractiveness of one morpheme automatically leads to lower rates of choosing the other morphemes. However, the question of whether outputs compete can still be addressed through reaction time and brain wave amplitude comparisons: if outputs do compete, it may take longer to respond in a nonce probe elicited production task when the neighborhood contains words bearing a number of different competitor morphemes than when it contains words bearing only one of the morphemes. Furthermore, there should be an increase in the amplitudes of the brain waves sensitive to morphological processing, such as the N400. If outputs do not compete, there should be no difference as long as the winning morpheme is controlled. This question remains for future research.

As tables 6.6-6.10 show, verbs containing -a- sound more natural when their roots end in /d/ than in /b/, when their roots end in /z/ than when they end in /ʒ/ but the latter still sound more natural than those whose roots end in /v/. Verbs containing -eva- sound more natural when their roots end in /m/ than in /n/, and in /d/ than in /b/. Finally, -i-bearing verbs sound more natural when their roots end in /z/ than when they end in /v/ or /ʒ/.

These findings contradict the hypothesis, proposed in Kapatsinski (2004b, c, d, 2005b), that co-occurrence-based associations between the place of articulation of the root-final consonant and the place of articulation of the suffix-initial vowel are responsible for the differences in productivity displayed by the stem extensions after different consonants. Rather, it appears that the units influencing -i-'s productivity must include the feature [continuant], those influencing -a-'s productivity must also be specified for voicing, at least in the case of stops, and those influencing -eva- must

specify nasality. Furthermore, the relevant vowel height must also be specified, since -eva-, a stem extension beginning with a [-high; -back] vowel and -i-, a stem extension consisting of a [+high; -back] vowel behave differently. Examining the behavior of -e- would help us determine whether the co-occurrence relations are phonological or morphological in nature, as would examining morphemes with the same phonological shape but a different function, e.g. the -i- and -a- noun plurals.

**Table 6.6. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /b/ vs. /d/ (reliable results are in bold)**

Extension → Root-final C ↓	-a-	-i-	-ova-	-eva-
/b/	<b>4.82</b>	5.17	4.32	<b>3.61</b>
/d/	<b>5.19</b>	5.23	4.53	<b>4.29</b>
Significance	<b>p=0.030</b>	p=0.754	p=0.595	<b>p=0.054</b>

**Table 6.7. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /p/ vs. /t/**

Extension → Root-final C ↓	-a-	-i-	-ova-	-eva-
/p/	4.94	4.94	4.62	4.36
/t/	4.95	5.29	4.22	4.72
Significance	p=0.947	p=0.210	p=0.157	p=0.307

**Table 6.8. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /m/ vs. /n/ (reliable results are in bold)**

Extension → Root-final C ↓	-a-	-i-	-ova-	-eva-
/m/	4.98	5.49	3.82	<b>4.78</b>
/n/	4.90	5.50	4.29	<b>4.12</b>
Significance	p=0.856	p=0.622	p=0.269	<b>p=0.040</b>

**Table 6.9. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva-: /v/ vs. /z/ vs. /ʒ/ (reliable results are in bold)**

Extension → Root-final C ↓	-a-	-i-
/v/	<b>4.50</b>	4.69
/z/	<b>5.23</b>	5.42
/ʒ/	<b>5.11</b>	4.67
Significance	<b>p=0.049</b>	p=0.071

**Table 6.10. Influence of the root-final consonant on the naturalness of verbs bearing -a-, -i-, -ova-, and -eva- (reliable results are in bold)**

Extension → Root-final C ↓	-a-	-i-
favorable voiced fricative(s)	/z/, /ʒ/: <b>5.19</b>	/z/: <b>5.42</b>
unfavorable voiced fricative(s)	/v/: <b>4.50</b>	/v/, /ʒ/: <b>4.68</b>
Significance	<b>p=0.033</b>	<b>p=0.019</b>

More work is needed to determine the exact nature of the unit, i.e. whether the associations are at the level of segments, syllabic constituents, or feature clusters. It is clear now that they are not at the level of individual features. Neither can they be at the level of the syllable because the syllables used in this study do not exist in the language<sup>44</sup>. It is also possible that roots and stem extensions are associated through units of various sizes. For instance, the feature [-palatalized] might be associated with the feature [+back], which might also be associated with [[+continuant][[Coronal][-anterior]]] or with [ʒ]. In the next chapter, we present an experiment investigating whether there are associations between the stem extensions and specific root-internal rimes and bodies. More work is needed to determine the cause of the associations, which could lie in articulation, perception, and/or co-occurrence. While the specifics remain to be worked out, it is clear that there are subsyllabic associations influencing naturalness ratings and morphological productivity.

<sup>44</sup> All root syllables are non-existent but phonotactically possible.

## **CHAPTER 7. EXPERIMENT III: RIMES AND BODIES MATTER<sup>45</sup>**

### ***7.1. Introduction***

Experiments I and II have shown that subsyllabic units within the root can develop associations with particular stem extensions or their parts. In this experiment, we examined whether the patterns of co-occurrence between the body and the rime of the root and particular stem extensions influences naturalness judgments in response to stimuli containing a syllabic constituent and either its associated stem extension or a stem extension not associated with it. The reasons rimes and bodies were chosen for testing were 1) rimes and bodies have been found to be psycholinguistically real (Yoon and Derwing 2001), 2) they are hypothesized to emerge from the patterns of segment co-occurrence (Kessler and Treiman 1997), 3) if both rimes and bodies are found to have an effect within a single language, this would provide support for the existence of overlapping units and thus for an exemplar model of memory (Bod 1992), and 4) body-based schemas for suffixes are inconsistent with the Rule-based Learner's similarity metric (Albright and Hayes 2002, 2003).

### ***7.2. Methods***

For every possible CV body and every possible VC rime in Russian, all verbs with monosyllabic roots containing the rime or body within the root and listed in the 125,000-word reverse dictionary of Russian (Sheveleva et al. 1974) were noted. For both rimes and bodies, three sets were created: constituents that overwhelmingly occurred with -i-,

---

<sup>45</sup> This chapter has been presented at CLS 31 (Kapatsinski 2005c).

those that overwhelmingly occurred with -a-, and those that did not occur in monosyllabic verbs. We also systematically varied the frequency of rimes and bodies within conditions where it was possible: bodies associated with -a- and rimes associated with either -a- or -i-. Nine stimulus categories were created: 1) -a-associated body, -a-associated rime, 2) -a-associated body, -i-associated rime, 3) -a-associated body, non-associated rime, 4) -i-associated body, -a-associated rime, 5) -i-associated body, -i-associated rime, 6) -i-associated body, non-associated rime, 7) non-associated body, -a-associated rime, 8) non-associated body, -i-associated rime, and 9) non-associated body, non-associated rime. For instance, /la/ occurs in 7 bisyllabic verbs bearing -a- (/labat<sup>i</sup>/, /lakat<sup>i</sup>/, /lajat<sup>i</sup>/, /latat<sup>i</sup>/, /laskat<sup>i</sup>/, /lajat<sup>i</sup>/, /lapat<sup>i</sup>/) and only 2 bisyllabic verbs bearing -i- (/ladit<sup>i</sup>/, /lazit<sup>i</sup>/) and is therefore labeled as co-occurring with -a-. On the other hand, /al/ occurs only in verbs bearing -i- (/valit<sup>i</sup>/, /ʒalit<sup>i</sup>/, /xvalit<sup>i</sup>/, /palit<sup>i</sup>/, /falit<sup>i</sup>/) and is therefore labeled as co-occurring with -i-. Therefore, in the stimulus root /lal/, the body co-occurs with -a- while the rime co-occurs with -i-. Co-occurrence statistics were also collected for the phonemes comprising the syllables from the same sources. The statistics were based on the form of the roots inside verbs rather than when they occurred independently as nouns. The reason this was done is that there is much support for the idea that speakers use product-oriented and not source-oriented generalizations (see Wang and Derwing 1994, Bybee 2001, chapter 8 of this thesis). Furthermore, doing this allows for more robust generalizations because there are more verbs than there are nouns which have served as bases for verbs.<sup>46</sup>

---

<sup>46</sup> Note that as long as generalizations are allowed to compete and are rewarded for having a high type frequency, as in the Rule-based Learner (Albright and Hayes 2003), product-oriented generalizations will outcompete source-oriented ones.

The task was the same as in experiment II. Thirteen adult native Russian speakers participated in the experiment. The stimuli were presented in a pseudorandomized order with no stimuli that had a body or rime favoring the same stem extension occurring in a row and no stimuli with the same rime or body occurring within five stimuli of each other. Half of the non-associated stimuli were stressed on the first syllable and half on the second. In unambiguously associated stimuli, stress was assigned to match the majority of the associates. With ambiguously associated stimuli, i.e. stimuli in which the body is associated with a different extension than the rime, two differently stressed variants were used if the -a-bearing associates and the -i-bearing associates differed in typical stress location. For analyzing the impact of the body, stimuli stressed where words sharing the body with it were stressed were used. For analyzing the impact of the rime, stimuli stressed where words sharing the rime with them were stressed were used. The stimuli are presented in table 7.1.

**Table 7.1. Stimuli used for testing whether rimes and bodies are associated with stem extensions they co-occur with**

Body co-occurs with	Rime co-occurs with			
	-a-		-i-	
-a-	map matʃ tim pim tit tiʃʃ tsoʃʃ tʃuv rit sit	mal mam lam lan lal lar tsoʃ tsotʃ tsol tsoz tsod	tsob ʒub ʒuz ʒud ʒuf tʃub tʃud tʃuf rib sib	maʃʃ laʃʃ tsof ʒuf tʃuf riŋ rif riʃʃ
-i-	ʃʃap ʃʃatʃ xofʃ dʃuv tuv	ʃʃal ʃʃam ʃʃjan ʃʃar xof xotʃ xoz xob dʃub	tud tul dʃuz dʃur dʃud dʃul dʃuf	ʃʃaʃ xof dʃuf tuf
Does not occur	tʃap gʃev tʃatʃ ʃʃev sʃap rʃofʃ sʃatʃ bʃuv dʃap vuv dʃatʃ zʃuv bʃap rʃuv bʃatʃ sʃuv xʃev git	tʃal bʃam tʃam bʃan tʃan bʃar sʃal gʃes sʃam xʃes sʃan ʃʃes sʃar bʃub dʃal vur dʃam zʃub dʃan rʃud dʃar sʃub bʃal gib	bʃam bʃan bʃar gʃes xʃes ʃʃes bʃub vur zʃub rʃud sʃub gib	tʃaʃʃ sʃaʃʃ dʃaʃʃ bʃaʃʃ rʃof bʃuf rʃuf sʃuf gin gʃif gʃiʃʃ

### **7.3. Results and discussion**

As table 7.2 shows, high frequency of the body increased the naturalness of the stimuli, regardless of whether the stimulus was bearing -a- or -i-. This result indicates that the naturalness judgments were based in part on the stimulus's probabilistic phonotactics.

More interesting is the finding that when the frequency of the body is controlled, verbs bearing -a- receive higher naturalness ratings if the body of their root co-occurs with -a- than if it co-occurs with -i-. Similarly, verbs bearing -i- receive higher ratings if their root's body co-occurs with -i- than when it co-occurs with -a-. This result indicates that root bodies can form associations with suffixes and therefore 1) adjacency is not necessary for association formation, and 2) it is necessary to take into account multi-segmental units that do not involve the root-final segment, and hence the Rule-based Learner's similarity metric (Albright and Hayes 2002, 2003) is incorrect. The difference is statistically significant when the results for -i- and -a- are combined and the body factor, coded as 'body favors the stem extension presented' with values Yes and No, is entered into the same ANOVA as respondent identity and rime's preference ( $p=0.022$ ). Just as in experiment II, the difference between bodies associated with -a- and bodies that do not occur in monosyllabic verbal roots was not significant for stimuli bearing -i- and vice versa ( $p=0.130$ ).

**Table 7.2. Body influence on naturalness judgments**

Stimulus bears → Body characteristics ↓		-a-	-i-
Frequency	Co-occurs with		
High	-a-	5.42	5.37
Low	-a-	5.27	4.85
Low	-i-	5.01	5.08
Low	Nothing	4.89	4.63

The results in Table 7.2 might be taken to imply that Russian is a body language, like Korean. However, Table 7.3 shows that this interpretation is incorrect because both the co-occurrence relations between root body and the suffix and those between root rime and the suffix play a role in naturalness judgments. For stimuli bearing -i-, rimes

frequently co-occurring with -i- increased the naturalness of the stimuli ( $p < 0.0005$ ). For stimuli bearing -a-, the difference was not reliable, suggesting that -a-rime associations are weaker, as predicted by the more even distribution of -a- across the lexicon. However non-associated rimes still produce significantly lower naturalness ratings ( $p = 0.001$ ). Comparing the results to the result in table 7.2, we can see that the abnormality in the results is the very high rating for -a-bearing stimuli whose rimes typically precede -i-. These rimes were more frequent in the experiment than rimes typically preceding -a- and thus the lack of significance may be due to a phonotactic confound. The finding that -i- is more sensitive to associations of the rime, combined with the finding that -a- and -i- are equally sensitive to whole-word similarity (Experiment II), suggests that when a nonce stimulus is presented, whole words are activated by the whole word representation of the stimulus, while associated sublexical representations are activated by sublexical parts of the stimulus.

**Table 7.3. Rime influence on naturalness judgments**

Stem extension presented → Rime associated with▼	-a-	-i-
-a-	5.25	4.76
-i-	5.21	5.31
None	5.01	4.67

Just as in experiment II, there were no differences between rimes associated with -a- and non-associated rimes for stimuli bearing -i- and vice versa.

It is also possible that the associations we investigate are really on the segmental level. However, neither the initial consonant, nor the vowel are reliable predictors of the differences between stimuli whose bodies favor -a- and those whose bodies favor -i-. The final consonant is a reliable predictor (in  $C_1VC_2$ ,  $p = .711$  for C1,  $p = .234$  for V,  $p = .039$  for

C2) but its predictions are in the wrong direction (the mean score is 4.97 for the stimuli it disfavors and 4.7 for those it favors). This indicates that the factor is simply significant by virtue of a negative correlation with a real determinant. This conclusion is supported by the fact that including C2 in the same model with body and rime reduces C2 to insignificance while body and rime stay significant ( $p=0.014$  for body,  $p=0.002$  for rime,  $p=0.171$  for C2). There is a significant interaction between C2, the body, and the rime ( $p<0.0005$ ), which explains why C2 has a significant but negative correlation. The differences among stimuli taking -i- predicted by rime's associations (Table 7.3) are not predicted by any of the segmental variables ( $p=0.309$  for C1,  $p=0.164$  for V,  $p=0.141$  for C2). Therefore, overlapping units are necessary to describe the data. The existence of overlapping units is predicted only by exemplar models of lexical representation (Bod 1992) and thus lends them strong support.

The existence of associations between root bodies and suffixes is also supported by findings reported by Ernestus and Baayen (2004) for Dutch where it was found that, although the choice of the past tense suffix is fully predictable from just the final consonant of the root, the rest of the root still plays a role in Dutch speakers' choice of the past tense suffix in nonce probe elicited production tasks. More work is needed to determine whether the effect is due to the presence of words favoring the suffix not favored by the root-final consonant in the stimulus's neighborhood (Ernestus and Baayen 2004) or associations between parts of the root other than the final consonant with the suffixes.

The similarity metric used by Albright and Hayes (2003) proposes that only units that involve the root-final segment of the root are relevant for suffix choice. The fact that

root bodies can develop associations with suffixes falsifies this claim. In chapter 8, we will see that it is possible to still use structured similarity without assuming that units associated with an affix must include a segment adjacent to it.

Finally, note that it is possible that the units involved are crosssyllabic. That is, the rimes are rimes if the syllable boundary in the verb falls at the root boundary but sequences of nucleus of the first syllable and the onset of the following syllable if onset maximization applies regardless of the root boundary. If the latter is the case, this finding argues against the notion of a syllable, which would be consistent with repeated failures of obtaining an effect of number of mismatched syllables in sound similarity judgments (Kapatsinski 2004e, chapter 3 of this thesis, Tamariz 2005). Especially problematic for the notion of a syllable is the finding that in chapter 3 we found an effect of number of matched and mismatched syllabic constituents and number of matched and mismatched segments while failing to find any effect of the number of matched and mismatched syllables. On the other hand, the relative weakness of the rime-based associations compared to body-based ones may be due to the ‘rimes’ really being units straddling a syllable boundary.

## CHAPTER 8. CONCLUSION

### *8.1. Summary*

Experiments I and II showed that high productivity does not necessarily lead to low sensitivity to the contents of a nonce stimulus's neighborhood. This finding provides evidence against the Dual Mechanism Model (Pinker and Prince 1988, 1994, Berent et al. 1999, Clahsen 1999, Pinker and Ullman 2002), which predicts that an inflectional domain has a default morpheme, which is applied via a categorical rule and is therefore the most productive morpheme with nonce stimuli similar to no existing words and is least sensitive to the phonological and semantic similarity relations between the stimulus and existing words.

Experiments II and III showed that stimuli similar to words taking a stem extension that is different from the stem extension presented are not less acceptable than stimuli similar to no existing words. This finding suggests that naturalness judgments are based on the number of similar verbs that have the stem extension presented and not on that number relative to the number of verbs bearing other stem extensions. If this hypothesis is confirmed in further studies, it would provide evidence against distributed connectionist models (e.g. Rumelhart and McClelland 1986). Since in connectionist models words are represented as overlapping activation patterns, a word is only as strong as its competitors are weak. Localist models, like LAST, do not make this prediction and can therefore accommodate lack of competition in naturalness ratings. Further evidence against connectionist models is provided by experiment B in chapter 3, which shows that sound similarity judgments are based on acoustically specific tokens, which are not

represented in standard connectionist models relying on a fixed set of input nodes.

Experiment III has shown that co-occurrence relations between the body of the root and the stem extension influence how natural the verb derived from the root with the extension sounds, and so do the co-occurrence relations between the stem extension and the root's rime. We have seen that the result cannot be predicted from segmental associations. The influence of the body is not predicted by the similarity metric proposed by Albright and Hayes (2003), since the body is irrelevant under the metric as long as the competing affixes are suffixes, as in the studies reported here.

Rather, it appears that, at least for words with monosyllabic roots, all parts of the word are taken into account to determine similarity. This is consistent with the findings from sound similarity judgments and suggests that phonological similarity may be assessed in the same way in tasks measuring phonological similarity explicitly and tasks that merely involve the use of similarity relations to perform some other operation.

With the directionality of Albright and Hayes' (2003) algorithm removed, the Rule-based Learner becomes indistinguishable from LAST because in LAST classical categories would also be predicted to be stronger than radial categories, as long as the potential member shares a necessary and sufficient attribute with members of the category and number of category members is controlled. The reason classical categories are expected to be able to attract new members more easily is that, once the part shared between the potential member and the current members is activated, all members of the classical category will get activated, swamping the members of competing radial categories, for which only some of the members would be activated at the same time.

Experiment III also provides support for exemplar models of lexical

representation by demonstrating the existence of partially overlapping sublexical phonological units. Thus, our results are fully consistent with a localist, exemplar-based implementation of Network Theory like LAST.

## ***8.2. Future directions***

### **8.2.1. Source-oriented vs. product-oriented schemas: Preliminary results**

Further work is needed to determine whether the generalizations discovered in this study are product oriented, source oriented, or both (cf. Wang and Derwing 1994, Bybee 2001). A priori, product-oriented generalizations are more likely due to the findings of Wang and Derwing (1994) for English. In addition, if generalizations are rewarded for having a high type frequency, as they are in the Rule-based Learner (Albright and Hayes 2003) and Network Theory (Bybee 1988, 1995), product-oriented generalizations will be more robust, at least for derivational morphology, since there are always more roots taking a suffix X than there are roots that can stand alone or occur with X and some other suffix.

Furthermore, the size of the unit on which type frequency calculations should be based remains a matter for future investigation. A particularly interesting phenomenon is suffix fusion, which occurs with -a- and both suggests that product-oriented generalizations are at work and provides a challenge for calculations of type frequency of -a- in certain phonological contexts.

A large number of -a-bearing verbs are verbs in which -a- is preceded by /nitʃ/, which follows the root of the verb, e.g. /plotnitʃati/ ‘to woodcarve’. Many of these words are derived from nouns bearing the agentive suffix -nik- (‘-er’), cf. /plotnik/ ‘woodcarver’. However, many of the verbs are derived from nouns that do not contain

-nik-, e.g. /nʲervnitʃatʲ/ ‘to be nervous’ is related to /nʲerv/ ‘nerve’ and /nʲervnʲij/ ‘nervous’. There is no word /\*nʲervnik/. The verb /kapriznitʃatʲ/ ‘to be capricious’ is related to /kapriz/ ‘caprice’ and /kapriznʲij/ ‘capricious’ but the word /\*kapriznik/ does not exist. The verb /xozʲajnitʃatʲ/ ‘to behave as if you are the owner’ is related to /xozʲain/ ‘owner’ but there is no word /\*xozʲajnik/. It thus appears that -nitʃatʲ has fused into a morpheme in these cases. Nonetheless, when the source word does contain -nik-, a second -nik- is not added, e.g. /\*plotnitʃnitʃatʲ/<sup>47</sup>. In addition, some words are stripped of their suffixes, which are replaced by /nitʃ/, e.g. /ʒulʲik/ ‘crook’ → /ʒulʲnitʃatʲ/ (/\*ʒulʲiknitʃatʲ/). The generalization is easy to formulate as a product-oriented generalization: a verb that means ‘to perform activities typical of X’ must end in -nitʃatʲ’ and have a monosyllabic root but is impossible to formulate in source-oriented terms.

The same case can be made for -ova- and -stvova- where -stv- is a suffix used to form property nouns. For instance, /ʃefstʲovatʲ/ ‘to act as a patron’ has been formed from /ʃefstʲovo/ ‘patronage’ (itself formed from /ʃef/ ‘boss, patron’), while /blagoʒelatelʲstvovatʲ/ ‘to wish well’ could not be formed from /\*blagoʒelatelʲstʲovo/, since the word does not exist: ‘well wishing’ is instead denoted by /blagoʒelatelʲnostʲ/ but /\*blagoʒelatelʲnostʲ{ext}tʲ/ is ungrammatical.<sup>48</sup>

<sup>47</sup> This restriction is morphological rather than phonotactic, e.g. /tʃernitʃnitʃatʲ/ ‘to gather blueberry’ is fine, since the first /nitʃ/ occurs across a morpheme boundary in side a strongly fused stem: tʃern- ‘black’ > /tʃernʲika/ ‘blueberry’ > /tʃernitʃnik/ ‘a place where much blueberry is growing’. Similarly /jedinitʃa/ ‘one’ > /edinitʃnʲij/ ‘single’ > /edinitʃnitʃatʲ/ ‘go at it alone’ (/\*edinitʃatʲ/). There also appears to be a preference for longer bases, e.g. /lesnitʃatʲ/ does not sound natural, even though /lesnik/ is a noun formed from /les/ ‘forest’ and meaning ‘forest guard’. These issues await systematic investigation.

<sup>48</sup> More controversially, the same analysis may be applied to -ova- and -irova-. For instance, /sovʲatʲ/ ‘to poke/thrust’ may have been formed by just attaching -a-, since all related nouns contain /sovʲ/ as the root, e.g. /zasovʲ/ ‘bolt’. In contrast, /pasovʲatʲ/ ‘to pass (in card games), to give up’ is clearly formed using -ova-

In the case of both /nitʃ/ and /stv/, there is a core of verbs that were formed from a noun or adjective through the addition of -a- and a number of more recent verbs that have been formed by adding a longer suffix. An interesting issue is whether words containing the longer suffixes also contain the shorter suffixes contained within the longer affixes for the speakers. In particular, do -ova-, -nitʃa-, -stvova-, -izova-, and -irizova- bearing types contribute to the type frequency of -a-, and do -stvova-, -izova-, and -irizova- bearing types contribute to the type frequency of -ova-. More generally, is type frequency based on the number of wordforms, bases, roots, lexemes, or morphemes adjacent to the morpheme whose type frequency is being calculated?

The fusion of these suffixes with each other and the concurrent rechunking of the first of the affixes to be more strongly attached to the second suffix than to the root is expected if subjects are relying on product-oriented generalizations: the frequency of the affix clusters is much greater than the frequency of any root+affix combination. Thus, there are 618 tokens of -nitʃ- followed by -a- in the Ogonek Corpus (SFB441, 2000)<sup>49</sup>. The most frequent root+nitʃ combination /trudnitʃ/ ‘laborer’ occurs 196 times, 41 of those tokens are followed by -a-. There are 1799 tokens of -stvova- in the Uppsala corpus<sup>50</sup> (Lonngren 1989), which is lower than only one root+suffix combinations:

---

from the noun /pas/ ‘pass’ (in card games). Similarly, /buksirovati/ ‘to tow’ is formed from /buksir/ and /pirovati/ ‘to feast’ is formed from /pir/ ‘feast’ using -ova- while /datirovati/ ‘to date (a document)’ is formed with -irova- from /data/ ‘date’ and /instrum<sup>l</sup>entirovati/ ‘arrange for (musical) instruments’ is formed from /instrum<sup>l</sup>ent/ ‘instrument’. A possible problem with this analysis, unlike with an analogous analysis for -stvova- and -nitʃa- is that there are very few verbs whose roots end in /ov/ and are followed by -a- and verbs whose roots end in /ir/ and are followed by -ova-, -ov- and -ir- are also not used as nominal derivational suffixes, at least productively since /ir/ is used in a few words, like /komandir/ ‘commander’, cf. /komanda/ ‘command’. Most of the nouns whose roots end in /ov/ are deverbal. Thus, there does not appear to be enough types for the merged suffix to become generalized and be applied to nouns not containing its first part.

<sup>49</sup> Determined by a Monoconc Pro search for \*?nichá?\*. Tokens include both verbs and deverbal nouns.

<sup>50</sup> Ogonek corpus could not be used because there are over 78000 tokens of -stv- in the corpus and files this large cannot be copied off the corpus site and into Monoconc.

/dejstv/ ‘do+stv’ with a frequency of 2285.<sup>51</sup> Hence, affix clusters are characterized by high frequency and high internal transitional probabilities. If, as we hypothesize, the fusion is the result of the high frequency of co-occurrence and transitional probability, more priming should be observed between words sharing the suffix cluster when the roots rarely co-occur with the first suffix.

The predictions of source-oriented and product-oriented generalizations also diverge for some root-final consonants. For instance, in the Sheveleva et al. (1974) dictionary most nouns ending in /z/ give rise to verbs bearing -a- while verbs whose roots end in /z/ are evenly split between -a- and -i-. In fact, the high productivity of -a- is a puzzle under a source-oriented model because most verbs bearing -a- do not have corresponding nouns. Thus in Sheveleva et al. (1974) 307 monosyllabic nouns form verbs using -i- and only 156 form verbs using -a-. On the other hand, only 598 bisyllabic verbs contain -i- while 722 contain -a-.

### **8.2.2. Schemas vs. analogy**

A related issue is whether the subjects are relying on schemas for the stem extensions or operating purely on the basis of analogy to words in the stimulus’s neighborhood. If the latter is the case, we would not expect any difference between -i-bearing and -a-bearing verbs in their sensitivity to similarity as long as the similarity difference they should display sensitivity to is controlled. Furthermore, if the former is the case, the influence of the token/type ratio should not depend on how evenly the tokens are distributed across

---

<sup>51</sup> While an affix in the cluster is very frequently followed by the next affix, i.e. has a reasonably high transitional probability, this factor does not predict suffix fusion because the root+suffix clusters are generally characterized by higher transitional probability. Thus, the probability of -a- given the occurrence of -nitf- is only 0.035, while the probability of -ova- given -stv- is only 0.073. By contrast, the probability of /stv/ given /dej/ is 0.465, given /chuv/ - 0.999. This is also a matter for future research.

types. Differences are expected if analogy is involved because of the non-linear nature of frequency effects: the same increase in token frequency of a stimulus has a greater effect if the stimulus's original token frequency is small. Thus, when the token/type ratio is controlled, we may expect less productivity from morphemes whose tokens are evenly distributed across types. On the other hand, if the effect of token/type ratio on productivity is due to the existence of a threshold beyond which a word no longer contributes to the type frequency of a morpheme, we may expect less even token/type distributions to decrease productivity at low morpheme token frequencies and the opposite to be true at high morpheme token frequencies since an even distribution of a morpheme with a low token frequency is likely to place no words above the threshold while an even distribution of a high frequency morpheme may put all words containing the morpheme above threshold.

### **8.2.3. Similarity metrics**

Another issue that awaits further investigation is the structure of the Russian syllable: better controls for rime frequency need to be introduced to determine whether -a- is in fact associated with some rimes more than with others. More work is also needed on the neighborhood radius and whether it is constant across tasks. The proper definition of a unit of similarity in Russian will have to rely on studies of sound similarity within Russian and checks of relevance of similarity ratings derived from sound similarity judgments for the measurement of morphologically-relevant phonological similarity

#### **8.2.4. Transitional probability versus constituent frequency**

Given our findings of overlapping sublexical units, it makes sense to ask whether noncategorical phonotactic effects are due to transitional probability between segments or features or to syllabic constituent frequency. While both frequency and transitional probability effects are predicted by LAST, a frequency effect or an effect of between-unit co-occurrence like the one found in Experiment III provides a clear indication that the units whose frequency subjects are sensitive to and whose co-occurrence statistics they monitor exist in their lexicons.

#### **8.2.5. LAST vs. connectionist models**

Finally, connectionist models and LAST can account for many of the same findings and yet make very different assumptions about lexical representations (distributed in the case of connectionist models, localist in the case of LAST). Therefore, it is important to be able to compare empirically the two models' predictions. Some further ways to test them are suggested below.

Hay's version of Network Theory specifically states that one can access words either directly or via their parts. Such an account contrasts with a full decomposition account (e.g. Stockall 2004) which states that words are always accessed through the morphemes they contain. The two accounts make different predictions for the effects of relative frequency on the speed of recognition of the derived form. Let us assume that connections between parts and wholes are trainable and that a connection is strengthened when the units it links are activated simultaneously. Then high-relative-frequency morphologically complex words are words that are very predictable given their base. In

other words, the connection leading from a base of a word of high relative frequency to the word itself is strong. Assuming that the strength of a connection influences either how fast activation can flow through it or how much activation leaving the node from which the connection originates will reach the node to which the connection leads, the listener will be able to activate the complex word from its base faster when the word is of high relative frequency. In a full decomposition model of lexical access, this will mean that words of high relative frequency will be recognized faster than words of low relative frequency. On the other hand, in a dual-route model, like that proposed by Hay (2001, 2003), high relative frequency increases the likelihood of accessing the word directly, thus assuming that the decompositional route is slowed down. For connectionist models a smaller part of a high-relative-frequency word is necessary for the network to react to the stimulus as it would react to the whole word, thus predicting, like full decomposition models, that high-relative-frequency words should be recognized faster. LAST is independent from models of segmentation and thus could be linked to either a full-decomposition or an optional-decomposition model. The predictions of the two accounts have yet to be tested.

In priming, LAST predicts that identity priming should have a shorter minimum prime-target stimulus onset asynchrony (SOA) at which it can be observed than similarity- or co-occurrence-based priming because activation needs time to spread from the prime type to the target type. Connectionist models, due to the use of distributed representation, predict that there should be no difference. Token frequency effects on priming are predicted to be due to prime frequency and target frequency in a localist

model like LAST but they are predicted to be due to the frequency of the features shared between the prime and the target in a connectionist model.

Age-of-acquisition effects are predicted to result from how well the unit is mastered in connectionist models (Steyvers and Tenenbaum 2005). Localist models predict that even equally familiar units with different ages of acquisition will still exhibit an even greater effect on connectivity because the familiar units acquired late get their familiarity from increased token frequency.

LAST predicts that there are no living beings for whom low-frequency stimuli prime or are primed less than high-frequency stimuli. Plaut and Booth (2000) hypothesize that this should be the case for subjects who are very slow at processing the stimuli.

Connectionist models assume that morphological similarity is just the result of phonological and semantic similarity and thus predict that morphological priming should be the same as priming between phonologically and semantically related pairs (Stockall 2004). LAST expects morphological priming to be stronger because root primes are lower in token frequency than phonological and (probably) semantic units and therefore spread more of their activation to the target. Thus, activation units reaching the target are larger when the prime is morphologically related to it through sharing the root. Since the larger the activation unit, the more slowly it decays, morphological priming is predicted to be not only greater in magnitude but also more long-lasting than phonological or semantic priming.

The Equity Principle, essential to LAST's functioning, assumes that strengthening an association between A and B will weaken priming between A and C, B and D, etc. (Zeelenberg 1998, Zeelenberg et al. 2003). The effect of direct and associative

activation of a stimulus is the same in connectionist models while only direct activation is predicted to lead to habituation in LAST, as found by Hall (2003). Finally, only LAST predicts that type frequency increases productivity while token/type ratio reduces it.

### ***8.3. Conclusion***

To conclude, this thesis has introduced a formalized version of Network Theory, presented evidence incompatible with the Dual Mechanism Model and has shown that the Rule-based Learner and Network Theory converge if the same, empirically supported similarity metric is used. Finally, we presented evidence problematic for connectionist models and identified ways in which they could be tested against the Local Activation Spread Theory introduced here.

## REFERENCES

- Aitchison, J. 1987. *Words in the mind: An introduction to the mental lexicon*. Oxford: Blackwell.
- Aitchison, J., and M. Straf. 1982. Lexical storage and retrieval: A developing skill. In A. Cutler, ed. *Slips of the tongue and language production*. Berlin: Mouton.
- Albright, A., and B. Hayes. 2002. Modeling English past tense intuitions with minimal generalization. In M. Maxwell, ed. *Proceedings of the 6<sup>th</sup> meeting of the ACL Special Interest Group in Computational Phonology*. Philadelphia: ACL.
- Albright, A., and B. Hayes. 2003. Rules vs. analogy in English past tenses: A computational/experimental study. *Cognition*, 90 (1), 119-61.
- Alegre, M., and P. Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40, 41-81.
- Anaki, D., and A. Henik. 2003. Is there a “strength effect” in automatic semantic priming? *Memory and Cognition*, 31.262-72.
- Anderson, J. R. 1976. *Language, memory, and thought*. Hillsdale: Lawrence Erlbaum.
- Anderson, J. R. 1983. *The architecture of cognition*. Cambridge: Harvard University Press.
- Anderson, J. R. 2000. *Cognitive psychology and its implications*. New York: Worth.
- Andrews, S. 1989. Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15.802-14.
- Arduino, L. S., and C. Burani. 2004. Neighborhood effects on nonword visual processing in a language with shallow orthography. *Journal of Psycholinguistic Research*, 33.75-95.
- Baayen, R., and R. Lieber. 1999. Productivity and English derivation: A corpus-based study. *Linguistics*, 29, 801-43.
- Baddeley, A., and B. Wilson. 1994. When implicit learning fails: Amnesia and the problem of error elimination. *Neuropsychologia*, 32.53-68.
- Bailey, T. M., and U. Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods. *Journal of Memory and Language*, 44, 568-91.
- Barnes, J. M., and B. J. Underwood. 1959. ‘Fate’ of first-list associations in transfer theory. *Journal of Experimental Psychology*, 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. *Journal of Memory and Language*, 44.350-75.

- Becker, C. A. 1979. Semantic context and word frequency effects in visual word recognition: An analysis of semantic strategies. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 252-9.
- Becker, S., M. Moscovitch, M. Behrmann, and S. Joordens. 1997. Long-term semantic priming: A computational account and empirical evidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23.1059-82.
- Beinert, R. A., and B. L. Derwing. 1993. Segment, rime, syllable, tier or root? Evidence from global sound similarity judgments in Arabic. *ESCOL '92*, 1-10.
- Bendrien, T. A. 1992. *Sound similarity judgments in English CVC's*. BA Honors Thesis: U of Alberta.
- Bentin, S., and L. B. Feldman. 1990. The contribution of morphological and semantic relatedness to the repetition effect at short and long lags: Evidence from Hebrew. *Quarterly Journal of Experimental Psychology*, 42A, 693-711.
- Bentin, S., and R. Frost. 1995. Morphological aspects of visual word processing in Hebrew. In L.B.Feldman, ed. *Morphological aspects of language processing*, 271-92. Hillsdale: Lawrence Erlbaum.
- Berent, I., S. Pinker, and J. Shimron. 1999. Default nominal inflection in Hebrew: Evidence for mental variables. *Cognition*, 72, 1-44.
- Berent, I., S. Pinker, and J. Shimron. 2002. The nature of regularity and irregularity: Evidence from Hebrew nominal inflection. *Journal of Psycholinguistic Research*, 31, 459-502.
- Berko, J. 1958. The child's learning of English morphology. *Word*, 14, 150-77.
- Bijeljac, R., A. Biarreau, and J. Grainger. 1997. Masked orthographic priming in bilingual word recognition. *Memory and Cognition*, 25.447-57.
- Blaxton, T. A. 1989. Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology*, 15, 657-68.
- Bod, R. 1992. A computational model of language performance: Data oriented parsing. *Proceedings of COLING '92*, Nantes, France.
- Bonin, P., M. Fayol, and M. Chalard. 2001. Age of acquisition and word frequency in written picture naming. *Quarterly Journal of Experimental Psychology*, 54A.469-89.
- Borowsky, R., and D. Besner. 1993. Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 813-40.
- Boudelaa, S., and M.G.Gaskell. 2002. A re-examination of the default system for Arabic plurals. *Language and cognitive processes*, 17, 321-43.
- Brew, C., and D. McKelvie. 1996. Word-pair extraction for lexicography. In K. Oflazer and H. Somers, eds. *Proceedings of the Second International Conference of New Methods in Language Processing*, 44-5.

- Browman, C. P. 1978. *Tip of the tongue and slip of the ear: Implications for language processing*. *UCLA Working Papers in Phonetics*, 42.
- Brown, G. D. A., and F. L. Watson. 1987. First in, first out: Word learning age and spoken word frequency as predictors of word familiarity and word naming latency. *Memory and Cognition*, 15, 208-16.
- Brown, R. 1973. *A first language: The early stages*. Cambridge: Harvard University Press.
- Brown, R., and D. McNeil. 1966. The “tip of the tongue” phenomenon. *Journal of Verbal Learning and Verbal Behaviour*, 5, 325-37.
- 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. 1985. *Morphology: a study of the relation between meaning and form*. Philadelphia: John Benjamins.
- Bybee, J. L. 1988. Morphology as lexical organization. In M. Hammond and M. Noonan, eds. *Theoretical morphology: Approaches in modern linguistics*, 119-41. San Diego: Academic Press.
- Bybee, J. L. 1995. Regular morphology and the lexicon. *Language and cognitive processes*, 10, 425-55.
- Bybee, J. L. 1999. Use impacts morphological representation. *Behavioral and brain sciences*, 22, 1016-7.
- Bybee, J. L. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow and S. Kemmer, eds. *Usage-based models of language*, 65-85. Stanford: CSLI.
- Bybee, J. L. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- Bybee, J. L., and C. L. Moder. 1983. Morphological classes as natural categories. *Language*, 59, 251-70.
- Bybee, J. L., and J. Newman. 1995. Are affixes more natural than stem changes? *Linguistics*, 33, 633-54.
- Bybee, J. L., W. Pagliuca, and R. Perkins. 1990. On the asymmetries in the affixation of grammatical material. In W. Croft, K. Denning, and S. Kemmer, eds. *Studies in diachronic typology for Joseph H. Greenberg*, 1-42. Amsterdam: John Benjamins.
- Bybee, J. L., and E. Pardo. 1981. On lexical and morphological conditioning of alternations: a nonce-probe experiment with Spanish verbs. *Linguistics*, 19, 937-68.
- Bybee, J. L., and D. I. Slobin. 1982. Rules and schemas in the development and use of the English past tense. *Language*, 58, 265-89.
- Carreiras, M., M. Perea, and J. Grainger. 1997. Effects of orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 857-71.

- Castles, A., C. Davis, and T. Letcher. 1999. Neighbourhood effects on masked form priming in developing readers. *Language and Cognitive Processes*, 14, 201-224.
- Cazden, C.B. 1968. The acquisition of noun and verb inflections. *Child Development*, 39, 433-48.
- Church, B., and D. L. Schachter. 1994. Perceptual specificity of auditory priming: Memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 521-33.
- Chwilla, D. J., P. Hagoort, C. M. Brown. 1998. The mechanism underlying backward priming in a lexical decision task: Spreading activation versus semantic matching. *Quarterly Journal of Experimental Psychology*, 51A, 531-60.
- Clahsen, H. 1999. Lexical entries and rules of language: A multidisciplinary study of German inflection. *Behavioral and Brain Sciences*, 22, 991-1060.
- Clahsen, H., M. Rothweiler, A. Woest, and G. F. Marcus. 1992. Regular and irregular inflection in the acquisition of German noun plurals. *Cognition*, 45, 225-55.
- Coleman, J., and J. B. Pierrehumbert. 1997. Stochastic phonological grammars and acceptability. *Computational Phonology* 3, 49-56. Somerset, NJ: ACL.
- Collins, A. M., and E. F. Loftus. 1975. A spreading activation theory of semantic memory. *Psychological Review*, 82, 407-28.
- Coltheart, M., E. Davelaar, J. T. Jonasson, and D. Besner. 1977. Access to the internal lexicon. In S. Dornic, ed. *Attention and performance VI*, 535-55. London: Academic Press.
- Coltheart, M., A. Woollams, S. Kinoshita, and C. Perry. 1999. A position-sensitive Stroop effect: Further evidence for a left-to-right component in print-to-speech conversion. *Psychonomic Bulletin and Review*, 6, 456-63.
- Connine, C. M., D. G. Blasko, and D. Titone. 1993. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 32: 193-210.
- Corina, D. P. 1992. Syllable priming and lexical representations: Evidence from experiments and simulations. In D. P. Corina (Ed.), *Proceedings of the 16th Annual Conference of the Cognitive Science Society* (pp. 779-784). Hillsdale, NJ: Erlbaum.
- Craik, F., and K. Kirsner. 1974. The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*, 26, 274-84.
- Crowder, R. G. 1973. Visual and auditory memory. In J. F. Kavanagh and I. G. Mattingly, eds. *Language by ear and by eye: The relationships between speech and reading*, 251-75. Cambridge: MIT Press.
- Cutler, A., and H.-C. Chen. 1995. Phonological similarity effects in Cantonese word recognition. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 1, 106-9. Stockholm, Sweden.
- Cutler, A., J. Hawkins, and G. Gilligan. 1985. The suffixing preference: A processing explanation. *Linguistics*, 23, 723-58.

- Cutler, A., B. van Ooijen, and D. Norris. 1999. Vowels, consonants and lexical activation. *Proceedings of the XIVth International Congress of Phonetic Sciences*, 3, 2053–6. San Francisco, CA.
- Dabrowska, E. 2001. Learning a morphological system without a default: the Polish genitive. *Journal of Child Language*, 28, 545-74.
- Derwing, B. L. 1992. A “pause-break” task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results from five diverse languages. *Language and Speech*, 35, 219-35.
- Derwing, B. L. 1987. A cross-linguistic experimental investigation of syllable structure. Part I: Background and methodology. *Proceedings of the Third Annual Meeting of the Pacific Linguistics Conference*, 93-102.
- Derwing, B. L., S. W. Cho, and H. S. Wang. 1991. A cross-linguistic experimental investigation of syllable structure: Some preliminary results. *Proceedings of the XIIth International Congress of Phonetic Sciences*, 3, 110-3.
- Derwing, B. L., M. L. Dow, and T. M. Nearey. 1987. Experimental evidence for phonological units. *Proceedings of the XIth International Congress of Phonetic Sciences*, 249-52.
- Derwing, B. L., and T. M. Nearey. 1995. On the status of the segment in four typologically diverse languages: Evidence from global sound similarity judgments. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 2, 598-601. Stockholm: KTH and Stockholm University.
- Derwing, B. L., and T. M. Nearey. 1986. Experimental phonology at the University of Alberta. In J. Ohala, ed. *Experimental phonology*, 187-209. San Diego: Academic Press.
- Derwing, B. L., D. B. Parkinson, and R. A. Beinert. 1995. Experimental investigations of Arabic syllable structure. *Perspectives on Arabic Linguistics* 7, 107-18.
- Derwing, B. L., and H. S. Wang. 1995. Concept formation as a tool for the investigation of phonological units. *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 3, 362-5.
- DiSciullo, A.-M., and E. Williams. 1987. *On the definition of word*. Cambridge: MIT Press.
- Dow, M. L. 1987. *On the psychological reality of sub-syllabic units*. PhD Thesis, U of Alberta.
- Dow, M. L. 1981. *On the role of orthography in experimental phonology*. MSc Thesis, U of Alberta.
- Draws, E., and P. Zwitserlood. 1995. Effects of morphological and orthographic similarity in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 21.1098-1116.
- Dumay, N., A. Benraiss, B. Barriol, C. Colin, M. Radeau, and M. Beson. 2001. Behavioral and electrophysiological study of phonological priming between

- bisyllabic spoken words. *Journal of Cognitive Neuroscience*, 13, 121-43.
- Emmorey, K. D. 1989. Auditory morphological priming in the lexicon. *Language and Cognitive Processes*, 4, 73-92.
- Emmorey, K., D. Corina, and U. Bellugi. 1995. Differential processing of topographic and referential functions of space. In K. Emmorey and J. Reilly, eds. *Language, gesture, and space*, 43-62. Hillsdale: Lawrence Erlbaum.
- Ervin, S. 1964. Imitation and structural change in children's language. In E. Lenneberg, ed. *New directions in the study of language*. Cambridge: MIT Press.
- Fallon, S. M., and D. C. Robinson. 1987. The effects of decorrelation of the discriminability of noise samples. *Journal of the Acoustical Society of America Supplement 1*, 81, S33.
- Fay, D., and A. Cutler. 1977. Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 8, 505-20.
- Feldman, L. B. 2003. What the repetition priming methodology reveals about morphological aspects of word recognition. In J. S. Bowers and C. J. Marsolek, Eds. *Rethinking implicit memory*, 124-38. New York: Oxford University Press.
- Forster, K. I., and C. Davis. 1991. The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language*, 30, 1-25.
- Forster, K. I. and C. Davis. 1984. Repetition priming and frequency attenuation in lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 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. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 696-713.
- Frauenfelder, U. H., R. H. Baayen, F. M. Hellwig, and R. Schreuder. 1993. Neighborhood density and frequency across languages and modalities. *Journal of Memory and Language*, 32, 781-804.
- Frisch, S. A., N. R. Large, and D. B. Pisoni. 2000. Perception of wordlikeness: Effects of segment probability and length on the processing of nonwords. *Journal of Memory and Language*, 42, 481-96.
- Fujimura, O., M. Macchi, and L.A. Streeter. 1978. Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech* 21, 337-46.
- Geiselman, R., and F. Belleza. 1977. Incidental retention of speaker's voice. *Memory and Cognition*, 5, 658-65.
- 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.

- Gibson, J. M., J. O. Brooks, L. Friedman, and J. A. Yesavage. 1993. Typography manipulations can affect priming of word stem completion in older and younger adults. *Psychology and Aging*, 8, 481-9.
- Glanzer, M., and S. L. Ehrenreich. 1979. Lexical access and lexical decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, 18.381-98.
- Godfrey, J. J., E. C. Holliman, and J. McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. *IEEE ICASSP*, 1517-20.
- Goldinger, S. D. 1992. *Words and voices: Implicit and explicit memory for spoken words*. Ph.D. Thesis: Indiana University, Bloomington.
- Goldinger, S. D.. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166-83.
- Goldinger, S. D.. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-79.
- Goldinger, S. D., T. Azuma, H. M. Kleider, and V. M. Holmes. 2003. Font-specific memory: More than meets the eye? In J. S. Bowers and C. J. Marsolek, Eds. *Rethinking implicit memory*, 157-96. New York: Oxford University Press.
- Goldinger, S. D., P. A. Luce, and D. B. Pisoni. 1989. Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language*, 28, 501-18.
- Goldinger, S. D., D. B. Pisoni, and J. S. Logan. 1991. On the nature of talker variability effects on serial recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 152-62.
- Gor, K., and T. Chernigovskaya. 2001. Rules in the processing of Russian verbal morphology. In *Current Issues in Formal Slavic Linguistics*, ed. by G. Zybatow, U. Junghanns, G. Mehlhorn, and L. Szucsich., 528-536. Frankfurt am Main: Vervuert Verlag.
- Gordon, B. 1983. Lexical Access and Lexical Decision: Mechanisms of frequency sensitivity. *Journal of Verbal Learning and Verbal Behavior*, 22.24-44.
- Gordon, P., and M. Alegre. 1999. Is there a dual system for regular inflections? *Brain and Language*, 68, 212-8.
- Graf, P., and L. Ryan. 1990. Transfer-appropriate processing for implicit and explicit memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16. 978-92.
- Greenberg, J. H., and J. J. Jenkins. 1964. Studies in the psychological correlates of the sound system of American English. *Word*, 20, 157-77.
- Guillaume, P. 1927/1973. The development of formal elements in the child's speech. In C.A.Ferguson and D.I.Slobin, eds. *Studies of child language development*, 240-51. New York: Holt, Rinehart, and Winston.

- Hall, C. J. 1992. *Morphology and mind: A unified approach to explanation in linguistics*. London: Routledge.
- Hall, G. 2003. Learned changes in the sensitivity of stimulus representations: Associative and nonassociative mechanisms. *Quarterly Journal of Experimental Psychology*, 56B.43-55.
- Halle, M. 1995. Feature geometry and feature spreading. *Linguistic Inquiry* 26: 1-46.
- Hare, M. L., J. L. Elman, and K. G. Daugherty. 1995. Default generalization in connectionist networks. *Language and Cognitive Processes*, 10, 601-630.
- Hare, M. L., N. Ford, and W. D. Marslen-Wilson. 2001. Frequency effects and the representation of regular verbs. In J. L. Bybee and P. Hopper, eds. *Frequency and the emergence of linguistic structure*, 181-200. Amsterdam: John Benjamins.
- Hay, J. 2001. Lexical frequency in morphology: Is everything relative? *Linguistics*, 39, 1041-70.
- Hay, J. 2003. *Causes and consequences of word structure*. London: Routledge.
- Hebb, D. O. 1949. *The organization of behavior*. New York: Wiley.
- Hintzman, D. L. 1986. 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, 105 (2), 251-79.
- Hintzman, D. L., and J. J. Summers. 1973. Long-term visual traces of visually presented words. *Bulletin of the Psychonomic Society*, 1, 325-7.
- Huntsman, L. A., and S. D. Lima. 2002. Orthographic neighbors and visual word recognition. *Journal of Psycholinguistic Research*, 31.289-306.
- Hurford, J. 1981. Malapropisms, left-to-right listing, and lexicalism. *Linguistic Inquiry*, 12, 419-23.
- Jacoby, L. L. 1983. Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 456-63.
- Jacoby, L. L., and M. Dallas. 1981. On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, 110, 306-40.
- Jacoby, L. L., and C. A. G. Hayman. 1987. Specific visual transfer in word identification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 456-63.
- Jaeger, J. J. 1980. Testing the psychological reality of phonemes. *Language and Speech*, 23, 233-53.
- Joanisse, M., and M. S. Seidenberg. 1999. Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences*. 96, 7592-7.
- Johnson, K. 1997. Speech perception without speaker normalization. In K. Johnson and J. W. Mullenix, eds. *Talker variability in speech processing*, 145-65. San Diego: Academic Press.
- Joordens, S., and S. Becker. 1997. The long and short of semantic priming effects in

- lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23.1083-1105.
- Jun, J. 1995. *A constraint based analysis of place assimilation typology*. PhD Thesis: UCLA.
- Kamin, L. J. 1969. Predictability, surprise, attention, and conditioning. In B. Campbell and R. Church, eds. *Punishment and aversive behavior*, 279–98. New York: Appleton-Century-Crofts.
- Kapatsinski, V. M. 2004. Measuring phonological similarity of English syllables. Ms. Available at <http://www.unm.edu/~alator/PhoneticSimilarity.pdf> (Forthcoming in proceedings of the *First UGA Linguistics Society Graduate Student Conference*).
- Kapatsinski, V. M. 2004b. Type frequency vs. phonotactic probability as determinants of Russian verbal stem extension productivity. 4<sup>th</sup> International Conference on the Mental Lexicon, Windsor, ON.
- Kapatsinski, V. M. 2004c. Frequency, neighborhood density, age-of-acquisition, and lexicon size effects in priming, recognition and associative learning: Towards a single-mechanism account. HDLS VI, Albuquerque, NM. (Forthcoming in the conference proceedings).
- Kapatsinski, V. M. 2004d. Defining attributes of the rule-based default are dissociable: Evidence contra the Dual Mechanism Model. HDLS VI, Albuquerque, NM.
- Kapatsinski, V. M. 2004e. Phonological similarity relations: Network organization of the lexicon and phonology. VIII Encuentro Internacional de Linguística en el Noroeste, Hermosillo, Sonora. (Forthcoming in the conference proceedings.)
- Kapatsinski, V. M. 2005. Characteristics of a rule-based default are dissociable: Evidence against the Dual Mechanism Model. In S. Franks, F. Y. Gladney, and M. Tasseva-Kurktchieva, eds. *Formal approaches to Slavic linguistics 13: The South Carolina meeting*, 136-46. Ann Arbor, MI: Michigan Slavic Publications.
- Kapatsinski, V. M. 2005b. To scheme or to rule: Defining attributes of the dual mechanism default are dissociable. BLS 31, Berkeley, CA.
- Kapatsinski, V. M. 2005c. Words as localist associative networks: Experimental evidence for an exemplar approach to the mental lexicon. CLS 41, Chicago, IL.
- Kessler, B., and R. Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37, 295-311.
- Kidd, G. R., and C. S. Watson. 1992. The “proportion-of-the-total-duration rule” for the discrimination of auditory patterns. *Journal of the Acoustical Society of America*, 92, 3109-18.
- Kim, J. J., D. J. Krupa, and R. F. Thompson. 1998. Inhibitory cerebello-olivary projections and the blocking effect in classical conditioning. *Science*, 279.570-3.
- Kim, J. J., S. Pinker, A. Prince, and S. Prasada. 1991. Why no mere mortal has ever flown out to center field. *Cognitive Science*, 15, 173-218.
- Kinoshita, S. 2000. The left-to-right nature of the masked onset effect in naming.

- Psychonomic Bulletin and Review*, 7, 133-41.
- Kinoshita, S. 2003. The nature of masked onset priming effects in naming: A review. The state of the art. In S. Kinoshita and S. J. Lupker, eds. *Masked priming*, 223-38. Hove: Psychology Press.
- Kiparsky, P. 1983. Lexical phonology and morphology. In I.S. Yang, ed. *Linguistics in the morning calm*, 3-91. Seoul: Hansin.
- Kirsner, K. 1973. An analysis of the visual component in recognition memory for verbal stimuli. *Memory and Cognition*, 1, 449-53.
- Kondrak, G. 2000. A new algorithm for the alignment of phonetic sequences. *Proceedings of the First Meeting of the North American Chapter of the Association for Computational Linguistics*, 228-95.
- Koepcke, K.-M. 1988. Schemas in German plural formation. *Lingua*, 74, 303-35.
- Koepcke, K.-M. 1998. The acquisition of plural marking in English and German revisited: Schemata versus rules. *Journal of Child Language*, 25.
- Koriat, A. 1981. Semantic facilitation in lexical decision as a function of prime-target association. *Memory and Cognition*, 9, 587-598.
- Koriat, A., and I. Lieblich. 1974. What does a person in a "TOT" state know that a person in a "don't know" state doesn't know. *Memory and Cognition*, 2, 647-55.
- Kuczaj, S. A. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16, 589-600.
- Kuczaj, S. A. 1981. More on children's initial failure to relate specific acquisitions. *Journal of Child Language*, 8, 485-7.
- Lachs, L., K. McMichael, and D. B. Pisoni. 2003. Speech perception and implicit memory: Evidence for detailed episodic encoding of phonetic events. In J. S. Bowers and C. J. Marsolek, Eds. *Rethinking implicit memory*, 215-38. New York: Oxford University Press.
- Ladefoged, P., and I. Maddieson 1986 *Some of the sounds of the world's languages*, UCLA Working Papers in Phonetics, 64.
- Lakoff, G. 1987. *Women, fire and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lobben, M. 1991. *Pluralization of Hausa nouns, Viewed from psycholinguistic experiments and child language data*. U of Oslo Master of Philosophy Thesis.
- Lonngren, L. 1989. The Uppsala corpus. Available at <http://heckel.sfb.uni-tuebingen.de/cgi-bin/cqp.pl>
- Luce, P. A., S. D. Goldinger, E. T. Auer, and M. S. Vitevitch. 2000. Phonetic priming, neighborhood activation, and PARSYN. *Perception and Psychophysics*, 62, 615-25.
- Luce, P. A., and D. B. Pisoni. 1998. Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1-36.

- Manso de Zuniga, C., G. Humphreys, and L. Evett. 1991. Additive and interactive effects of repetition, degradation, and word frequency in the reading of handwriting. In D. Besner and G. Humphreys, eds. *Basic processes in reading*, 10-33. Hillsdale, NJ: Erlbaum.
- Marchant, H. G., and J. W. Moore. 1973. Blocking of the rabbit's conditioned nictitating membrane response in Kamin's two-stage paradigm. *Journal of Experimental Psychology*, 101.155-8.
- Marcus, G. F. 1998. *The algebraic mind: Integrating connectionism and cognitive science*. Cambridge: MIT Press.
- Marcus, G. F., S. Pinker, M. Ullman, J. Hollander, T. J. Rosen, and F. Xu. 1992. Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, 57, Serial No. 228.
- Marslen-Wilson, W. 1987. Functional parallelism in spoken word-recognition. *Cognition* 8,1-71.
- Marslen-Wilson, W. 1990/1995. Activation, competition, and frequency in lexical access. In: Altmann, G. T. M., ed. *Cognitive models of speech processing: psycholinguistic and computational perspectives*, 148-73. Cambridge, MA: MIT.
- McCarthy, J., and A. Prince. 1990. Foot and word in prosodic morphology: The Arabic broken plural. *Natural Language and Linguistic Theory*, 8, 209-83.
- McClelland, J., and K. Patterson. 2002. Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Sciences*, 6 (11), 465-72.
- 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. *Psychological Review*, 88.375-407.
- McGovern, J. B. 1964. Extinction of associations in four transfer paradigms. *Psychological Monographs*, 78.
- McInish, J. R., and R. S. Tikofsky. 1969. Distinctive features and response latency: A pilot study. *Perception and Psychophysics*, 6, 267-8.
- McLennan, C., P. A. Luce, and J. Charles-Luce. 2003. Representation of lexical form in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 539-53.
- Melamed, I. D. 1999. Bitext maps and alignment via pattern recognition. *Computational Linguistics*, 25(1), 107-30.
- Meschyan, G., and A. Hernandez. 2002. Age-of-acquisition and word frequency: Determinants of object-naming speed and accuracy. *Memory and Cognition*, 30.262-9.
- 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, G. A., and P. Nicely. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338-52.

- 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. Thesis: SUNY Buffalo.
- Monsell, S. 1991. The nature and locus of word frequency effects in reading. In *Basic processes in reading: Visual word recognition*, ed. by D. Besner and G.W. Humphreys. Hillsdale, Hove, and London: Lawrence Erlbaum.
- Morford, J. 2003. Why does exposure to language matter? In T. Givon and B. F. Malle, eds. *The evolution of language out of pre-language*, 329-41. Amsterdam: John Benjamins.
- Morrison, C. M., K. W. Hirsh, T. Chappell, and A. W. Ellis. 2002. Age and age of acquisition: An evaluation of the cumulative frequency hypothesis. *The European Journal of Cognitive Psychology*, 14.435-59.
- Moscato 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.
- Mueller, S. T., T. Seymour, A. Krawitz, D. A. Kieras, and D. E. Meyer. 2003. Theoretical implications of articulatory duration, phonological similarity and phonological complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 29, 1353-80.
- 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 D. Besner and G. W. Humphreys, eds. *Basic processes in reading: Visual word recognition*. Hillsdale, Hove, and London: Lawrence Erlbaum.
- Nelson, D. L., and L. D. Nelson. 1970. Rated acoustic (articulatory) similarity for word pairs varying in number and ordinal position of common letters. *Psychonomic Science*, 19, 81-2.
- Nevers, B., and R. Versace. 1998. Knowledge acquisition in long-term memory: Activation and construction of traces. In G. Ritschard, A. Berchtold, F. Duc, and A. D. Zighed, eds. *Apprentissage: Des principes naturels aux methods artificielles*. Paris: Hermes.
- 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.
- Newman, R. S., J. R. Sawusch, and P. A. Luce. 1997. Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 873– 89.
- Ng, K. 1999 Towards robust methods for spoken document retrieval. In *Proceedings of the International Conference on Spoken Language Processing*, 3, 939-42.

- Norris, D. G. 1984. The effects of frequency, repetition and stimulus quality in visual word recognition. *Quarterly Journal of Experimental Psychology*, 36A, 507-18.
- Nosofsky, R.M. 1988. Similarity, frequency and category representation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 54-65.
- Nosofsky, R. M. 1990. Relations between exemplar similarity and likelihood models of classification. *Journal of Mathematical Psychology* 34: 393-418.
- Ohala, J. 1990. The phonetics and phonology of aspects of assimilation. In J. Kingston and M. Beckman, eds. *Papers in Laboratory Phonology I: Between the grammar and the physics of speech*, 258-75. Cambridge University Press, Cambridge.
- Paap, K. R. J. E. McDonald, R. W. Schvaneveldt, and R. W. Noel. 1987. Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart, ed. *Attention and performance XII*. Hove: Lawrence Erlbaum.
- Palmeri, T. J., S. D. Goldinger, and D. B. Pisoni. 1993. Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 309-28.
- 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, 668-677.
- Peresotti, F., and J. Grainger. 1999. The role of letter identity and letter position in orthographic priming. *Perception and Psychophysics*, 61, 691-706.
- 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 J.L.Bybee and P.Hopper, eds. *Frequency and the emergence of linguistic structure*, 137-57. Amsterdam: John Benjamins.
- Pinker, S. 1999. *Words and rules: The ingredients of language*. New York: Basic Books.
- Pinker, S., and A. Prince. 1988. On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition*, 28, 73-193.
- Pinker, S, and A. Prince. 1994. Regular and irregular morphology and the psychological status of rules of grammar. In S.D.Lima, R.L.Corrigan, and G.K.Iverson, eds. *The reality of linguistic rules*, 321-51. Amsterdam, Philadelphia: John Benjamins.
- Pinker, S., and M. T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences*, 6 (11), 456-63.
- 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. *Psychological Review*, 107, 786-823.
- Plaut, D. C., J. L. McClelland, M. S. Seidenberg, and K. Patterson. 1996. Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103, 56-115.

- Praamstra, P., A. S. Meyer, and W. J. M. Levelt. 1994. Neurophysiological manifestations of phonological processing: Latency variation of a negative ERP component time-locked to phonological mismatch. *Journal of Cognitive Neuroscience*, 6, 204–19.
- Praamstra, P., and D. F. Stegeman. 1993. Phonological effects on the auditory N400 event-related brain potential. *Cognitive Brain Research*, 1, 73–86.
- Prasada, S., and S. Pinker. 1993. Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes*, 8, 1-56.
- Radeau, M. 1995. Facilitatory and inhibitory phonological priming effects as a function of overlap position. *Abstracts of the 36th Annual Meeting of the Psychonomic Society*, 59.
- Radeau, M., J. Morais, and J. Segui. 1995. Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 1297-1311.
- Radeau, M., J. Segui, and J. Morais. 1994. The effect of overlap position in phonological priming between spoken words. *Proceedings of the International Conference on Spoken Language Processing*, 1419–22.
- Ramscar, M. 2002. The role of meaning in inflection: Why the past tense does not require a rule. *Cognitive Psychology*, 45, 45-94.
- Ratcliff, R., and G. McKoon. 1981. Does activation really spread? *Psychological Review*, 88.454-62.
- Ratcliff, R., and G. McKoon. 1988. A retrieval theory of priming in memory. *Psychological Review*, 95, 383-408.
- Roeper, T. 1999. Universal bilingualism. *Bilingualism, Language and Cognition*, 2, 169-86
- Rubin, D. C. 1975. Within word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior*, 14, 392-7.
- Rueckl, J. G. 2003. A connectionist perspective on repetition priming. In J. S. Bowers and C. J. Marsolek, Eds. *Rethinking implicit memory*, 67-104. New York: Oxford University Press.
- Rumelhart, D.E., and J.L.McClelland. 1986. On learning the past tenses of English verbs. In D.E.Rumelhart and J.L.McClelland, eds. *Parallel distributed processing: Vol.2: Psychological and biological models*. Cambridge: MIT Press.
- Ryle, G. 1949. *The concept of mind*. London: Hutchinson.
- Salasoo, A., and D. B. Pisoni. 1984. Sources of knowledge in spoken word recognition. *Journal of Verbal Learning and Verbal Behavior*, 23, 210-31.
- Savage, C., Lieven, E., Theakston, A., & Tomasello, M. (2003). Testing the abstractness of young children's linguistic representations: Lexical and structural priming of syntactic constructions? *Developmental Science*, 6-5, 557-567.

- Scarborough, D.L., C. Cortese, and H.S. Scarborough. 1977. Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 1-17.
- Schachter, D. L., and B. Church. 1992. Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 915-30.
- Schiller, N. O. 1998. The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, 39, 484-507.
- Schiller, N. O. 2004. The onset effect in word naming. *Journal of Memory and Language*, 50, 477-90.
- Schriefers, H., A. Frederici, and P. Graetz. 1992. Inflectional and derivational morphology in the mental lexicon: Symmetries and asymmetries in repetition priming. *Quarterly Journal of Experimental Psychology*, 44A.373-90.
- Schuberth, R. E., and P. Eimas. 1977. Effects of context on the classification of words and non-words. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 27-36.
- Schuberth, R. E., K. T. Spoehr, and D. M. Lane. 1981. Effects of stimulus and contextual information on the lexical decision process. *Memory and Cognition* 9, 68-77.
- Schvaneveldt, R. W., and J. E. McDonald. 1981. Semantic context and the encoding of words: Evidence for two models of stimulus analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 7.673-87.
- Schwantes, F. M. 1985. Expectancy, integration, and interactional processes: Age differences in the nature of words affected by sentence context. *Journal of Experimental Child Psychology*, 39.212-29.
- Sears, C. R., S. J. Lupker, and Y. Hino. 1999. Orthographic neighborhood effects in perceptual identification and semantic categorization tasks: A Test of the multiple readout model. *Perception and Psychophysics*, 61.1537-54.
- Segui, J., and J. Grainger. 1990. Priming word recognition with orthographic priming: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance*, 16.65-76.
- Seidenberg, M. S., and J. L. McClelland. 1989. A distributed, developmental model of word recognition and naming. *Psychological Review*, 96, 523-68.
- Selkirk, E. O. 1982. The syllable. In H. van der Hulst and N. Smith, eds. *The structure of phonological representations (Part II)*, 337-84. Dordrecht: Foris.
- Sereno, J. A., and A. Jongman. 1997. Processing of English inflectional morphology. *Memory and Cognition*, 24, 425-37.
- SFB 441, Project B1. 2000. *Ogonek 1996-2000*. <http://heckel.sfb.uni-tuebingen.de/cgi-bin/cqp.pl>
- Shvedova, N. Ju., N. D. Arutjunova, A. V. Bondarko, V. V. Ivanov, V. V. Lopatin, I. S. Uluxanov, F. I. Filin, Ju. S. Maslov, and D. N. Shmelev. 1980. *Russkaja grammatika*.

Vol. 1. Moscow: Nauka.

- Simpson, G. B., and T. C. Lorsch. 1983. The Development of automatic and conscious components of contextual facilitation. *Child Development*, 54, 760-72.
- Skousen, R. 1989. *Analogical modeling of language*. Dordrecht: Kluwer.
- Slobin, D. I. 1973. Cognitive prerequisites for the development of grammar. In C. A. Ferguson and D. I. Slobin, eds. *Studies of Child Language Development*, 177-208. New York: Holt, Rinehart, and Winston.
- Slowiacek, L. M., and M. Hamburger. 1992. Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1239-50.
- Slowiacek, L. M., J. M. McQueen, E. Soltano, and M. Lynch. 2000. Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language*, 43, 530-60.
- Slowiacek, L. M., H. C. Nusbaum, and D. B. Pisoni. 1987. Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 13, 64-75.
- Sommers, M. S., and B. P. Lewis. 1999. Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language*, 40, 83-108.
- Stanovich, K. E., and R. F. West. 1981. The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 658-72.
- Stanovich, K. E., and R. F. West. 1983. On priming by a sentence context. *Journal of Experimental Psychology: General*, 112, 1-36.
- Stanovich, K. E., R. F. West, and D. Feenan. 1981. A longitudinal study of sentence-context effects in second-grade children: Tests of an interactive-compensatory model. *Journal of Experimental Child Psychology*, 32, 185-99.
- Stark, C. E. L. 1997. *Repetition priming of words, pseudowords, and nonwords*. Ph.D. Thesis: Carnegie Mellon University.
- Stark, C. E. L., and J. L. McClelland. 2000. Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 945-972.
- Stemberger, J. P. 1981. Morphological haplology. *Language*, 57, 791-817.
- Stemberger, J. P., and B. MacWhinney. 1986. Frequency and lexical storage of regularly inflected forms. *Memory and Cognition*, 14, 17-26.
- Stemberger, J. P., and B. MacWhinney. 1988. Are inflected forms stored in the lexicon? In M. Hammond and M. Noonan, eds. *Theoretical morphology: Approaches in modern linguistics*, 101-16. San Diego: Academic Press.
- Steriade, D. 2001. Directional asymmetries in place assimilation. In E. Hume and K. Johnson, eds. *The role of speech perception in phonology*. New York: Academic

- Press.
- Steriade, D. 2004. Knowledge of similarity and narrow lexical override. *BLS* 29, 583-98.
- Steriade, D., and J. Zhang. 2001. Context-dependent similarity. Paper presented at CLS 33.
- 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. *Magentoencephalographic [sic] investigations of morphological irregularity and identity*. Ph.D. Thesis, MIT.
- Storkel, H. L. 2001. Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, Hearing Research* 44, 1321-37.
- Storkel, H. L. 2004. Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, 25, 201-21.
- 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. Thesis, UCLA.
- Tamariz, M. 2005. *Exploring the adaptive structure of the mental lexicon*. Ph.D. Thesis, U of Edinburgh.
- 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.
- Titone, D., and C. M. Connine. 1997. Syllabification strategies in spoken word processing: Evidence from phonological priming. *Psychological Research*, 60, 251-63.
- Treiman, R. 1983. The structure of spoken syllables: Evidence from novel word games. *Cognition*, 15, 49-74.
- Treiman, R. 1984. On the status of final consonant clusters in English syllables. *Journal of Verbal Learning and Verbal Behavior*, 23, 343-56.
- Treiman, R. 1985. Onsets and rimes as units of spoken syllables: Evidence from children. *Journal of Experimental Child Psychology*, 39, 161-81.
- Treiman, R. 1988. Distributional constraints and syllable structure in English. *Journal of Phonetics*, 16, 221-9.
- Treiman, R., and C. Danis. 1988. Short-term memory errors for spoken syllables are affected by the linguistic structure of the syllables. *Journal of Experimental*

- Psychology: Learning, Memory, and Cognition*, 14, 145-52.
- Treiman, R., and B. Kessler. 1995. In defense of an onset-rime syllable structure for English. *Language and Speech*, 38, 127-42.
- Tweney, R., S. Tkacz, and S. Zaruba. 1975. Slips of the tongue and lexical storage. *Language and Speech*, 18, 388-96.
- Ullman, M. 1999. Acceptability ratings for regular and irregular past tense forms: Evidence for a dual-system model of language from word frequency and morphological neighborhood effects. *Language and Cognitive Processes*, 14, 47-67.
- Van Kampen, J. 1997. *First steps in wh-movement*. Ph.D. Thesis: Utrecht University.
- Versace, R. 1998. Frequency and prime duration effects on repetition priming and orthographic priming with words and pseudowords. *Cahiers de Psychologie Cognitive*, 17, 535-56.
- 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.
- Vitz, P. C., and B. S. Winkler. 1973. Predicting the judged "similarity of sound" of English words. *Journal of Verbal Learning and Verbal Behavior*, 12, 373-88.
- Vroomen, J., and B. de Gelder. 1997. Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 710-20.
- Wang, H. S., and B. L. Derwing. 1993. Is Taiwanese a "body" language? *Proceedings of the 1993 Annual Conference of the Canadian Linguistic Association*, 679-94.
- Wang, H. S., and B. L. Derwing. 1994. Some vowel schemas in three English morphological classes: Experimental evidence. In M. Y. Chen and O. C. L. Tseng, eds. *In honor of Professor William S.-Y. Wang: Interdisciplinary studies on language and language change*, 561-75. Taipei: Pyramid Press.
- West, R. F., and K. E. Stanovich. 1982. Source of inhibition in experiments on the effect of sentence context on word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 385-99.
- Whittlesea, B. W. A. 2003. On the construction of behavior and personal experience: The production and evaluation of performance. In J. S. Bowers and C. J. Marsolek, Eds. *Rethinking implicit memory*, 259-60. New York: Oxford University Press.
- Williams, E. 1981. On the notions 'lexically related' and 'head of word'. *Linguistic Inquiry*, 12, 245-74.
- Wiebe, G. E., and B. L. Derwing. 1994. A forced-choice word-blending task for testing intra-syllabic breakpoints in English, Korean, and Taiwanese. *The Twenty-first Lacus Forum*, 142-52.

- Wilson, B., A. Baddeley, and J. Evans. 1994. Errorless learning in the rehabilitation of memory-impaired people. *Neuropsychological Rehabilitation*, 3.307-26.
- Wilson, B., and J. Evans. 1996. Error-free learning in the rehabilitation of people with memory impairments. *Journal of Head Trauma Rehabilitation*, 11.54-64.
- Yoon, Y. B. 1994. Syllable saliency in Korean: Evidence from the sound similarity judgment task. *University of Alberta Papers in Experimental and Theoretical Linguistics*, 76-87.
- Yoon, Y. B. 1995. *Experimental studies of the syllable and the segment in Korean*. PhD Thesis: U of Alberta.
- Yoon, Y. B., and B. L. Derwing. 2001. A language without a rhyme: Syllable structure experiments in Korean. *Canadian Journal of Linguistics*, 46 (3/4), 187-237.
- Zaliznjak, A. A. 1977. *Grammaticheskij slovar' russkogo jazyka: Slovoizmenenie*. Moscow: Russkij Jazyk.
- Zeelenberg, R. 1998. *Testing Theories of Priming*. Ph. D. Dissertation: U of Amsterdam.
- Zeelenberg, R., D. Pecher, and J. G. W. Raaijmakers. 2003. Associative repetition priming: A selective review and theoretical implications. In J. S. Bowers and C. J. Marsolek, eds. *Rethinking implicit memory*, 261-83. New York: Oxford University Press.
- Zevin, J.D., and M.S.Seidenberg. 2002. Age of acquisition effects in word reading and other tasks. *Journal of Memory and Language*, 47, 1-29.
- 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*, 32.31-8.
- Ziegler, J. C., M. Muneaux, and J. Grainger. 2003. Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language*, 48, 779-93.
- Zimmer, K.E. 1969. Psychological correlates of some Turkish morpheme structure conditions. *Language* 45: 309-321.