Aphasic Speech Errors:

Spontaneous and Elicited Contexts

Jean K. Gordon

School of Communication Sciences and Disorders

McGill University, Montréal, Québec

August, 2000

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Ph.D.

0-612-70032-1

Table of Contents

List of Figures

List of Appendices

Acknowledgements

I gratefully acknowledge the generous funding of the Natural Sciences and Engineering Research Council of Canada, and the Fonds de la Recherche en Santé du Québec, which together made this research possible.

I am also heavily indebted to the many people who helped to ensure the completion of this project. My gratitude to all those who provided their materials, their time, and their expertise to facilitate the research:

- first and foremost, to the people who selflessly agreed to be my subjects;
- to Deborah Gagnon and Joseph Marin for providing PNT materials, and to Adelyn Brecher and Nadine Martin for answering endless questions about error coding;
- to David Pisoni, Paul Luce, and Mike Vitevitch for providing access to the on-line lexical database, and to Mike Vitevitch again for explaining it;
- to The Curtis Publishing Company for granting me permission to reproduce the Norman Rockwell pictures;
- to Fabrice Rouah and James Ramsay for statistical consultation;
- to Elin Thordardottir and André-Yves Gagnon for help with the French translation;
- and to Nancy Azevedo, Kathy Palesch, Anita Shuper, Emma Duerden, Katrine Doucet, and Wendi Aasland for their patient assistance with the testing, transcription, and scoring.

Thanks also to all the faculty and Ph.D. students at the School of Communication Sciences and Disorders for their support and encouragement. Special thanks to my committee members—Marc Pell and Hugh Buckingham—for providing insightful and helpful comments, and for doing so under completely unrealistic time constraints. Most importantly, I am eternally grateful for the wisdom, patience, and generosity of my supervisor, Shari Baum, who has provided moral and intellectual support on demand for the past five years.

My appreciation also goes to friends and family: thanks to Judie for being in France and not distracting me with late-night games of backgammon; thanks to Colin for doing all this before me and telling me what he learned; thanks to my parents and my brothers for their understanding and encouragement. This thesis is dedicated to my father, whose intellectual curiosity, perseverance and strength of character have always been an inspiration to me, but especially in the final weeks of his life. Thanks, Dad.

Abstract

The goal of the current study was to investigate the retrieval of phonological word forms during the speech production of persons with aphasia, in order to inform models of the structure and function of the phonological lexicon. Using a naturalistic, connected speech task (picture description) and a more structured, single-word production task (picture naming) several characteristics of the target and its phonological 'neighbourhood' were examined, specifically: the target word's frequency of occurrence; the number of words which are phonologically similar to the target (neighbourhood density); and the average frequency of those 'neighbours' (neighbourhood frequency).

To assess the influence of these factors on a *target's susceptibility to error*, the neighbourhood values of the words produced incorrectly in the picture description task were compared to those of a comparable corpus of correctly produced words from the same speech samples. In the naming task, target susceptibility was assessed by analyzing the error rates on individual stimulus items. The results of both tasks indicated that the lower a target's frequency of occurrence was, and the fewer neighbours it had, the more susceptible it was to error. To assess the impact of the neighbourhood on *the outcome of the error*, neighbourhood values of the errors produced were compared to those of their targets. In neither task were errors found to differ significantly from their targets in frequency or neighbourhood density.

These results contribute to the literature on lexical access primarily by extending findings of neighbourhood effects in normal speech production to the aphasic population. In doing so, the present study lends support to the basic tenets of the Neighborhood Activation Model (Luce & Pisoni, 1998), and to the notion of the continuity thesis, in which aphasic deficits are hypothesized to reflect quantitative,

rather than qualitative, differences from normal processing. Results are also in agreement with previous studies illustrating that aphasic error outcomes are strongly constrained by a number of linguistic factors which also constrain normal error production. Results are interpreted as consistent with an interactive connectionist framework of speech production.

Resumé

Le but de cette étude était d'examiner l'accès aux formes phonologiques des mots pendant la production de la parole par des personnes atteintes d'aphasie, afin d'informer des modèles du lexique phonologique. Au moyen de deux tâches—l'une plus naturelle et spontanée (description d'une scène) et l'autre plus structurée (dénomination d'un objet)—plusieurs caractéristiques du mot cible et son 'voisinage' phonologique ont été examinés, tels: la fréquence d'occurrences de la cible; le nombre de mots phonologiquement semblables à la cible (densité de voisinage); et la fréquence moyenne de ces 'voisins'.

Pour évaluer l'influence de ces facteurs sur *la sensibilité des cibles à l'erreur*, les valeurs de voisinage des mots produits inexactement dans la tâche de description d'image ont été comparées aux valeurs d'un corpus de mots produits correctement dans les mêmes échantillons de parole. Dans la tâche de dénomination, la sensibilité de la cible a été évaluée en analysant le taux d'erreurs des stimuli individuels. Les retombées des deux tâches ont indiqué que les cibles les plus sensibles à l'erreur étaient celles qui étaient les moins fréquentes et qui avaient le moins de voisins. Pour évaluer l'impact du voisinage sur *la nature des erreurs*, les valeurs de voisinage des erreurs produites ont été comparées à celles de leurs cibles. Dans aucune de ces tâches a-t-on retrouvé d'importantes différences entre les erreurs et leurs cibles, en termes de fréquence ou densité de voisinage.

Ces résultats contribuent aux écrits sur l'accès lexical en étendant à la population aphasique les effets de voisinage en production normale. De cette manière, la présente étude appuie les principes de base du 'Neighborhood Activation Model' (Luce et Pisoni, 1998), et la notion de la 'thèse de continuité', dans laquelle les déficits aphasiques reflètent des différences quantitatives, plutôt que qualitatives, par

comparaison au traitement normal. De plus, les résultats sont en accord avec des études précédentes illustrant que la nature des erreurs aphasiques est fortement contrainte par plusieurs facteurs linguistiques qui contraignent également la production normale. Les résultats sont interprétés comme étant compatibles avec un modèle interactif connectioniste de production de la parole.

Introduction. The Context of Aphasic Speech Errors

One of the most compelling aspects of aphasia is the production of paraphasic utterances. Hesitations, false starts, and vague referents are clear indications of word-finding difficulties, in both normal and aphasic language production, but the inaccurate and even nonsensical words or non-words that are sometimes produced by aphasic speakers in place of an intended utterance are jarring reminders of the pathological workings of the damaged brain. Researchers hope that their analysis might provide a glimpse into the often circuitous and misdirected route through which the mind afflicted with aphasia travels in search of the correct lexical item. Of theoretical interest, paraphasic errors demonstrate striking similarities to the slips of the tongue produced in the course of normal conversation, yet can also appear as one of the most bizarre manifestations of aphasic production. Paraphasias are also of clinical importance, because they suggest both the nature of the lexical access deficit to be addressed in therapy and the types of strategies which might be useful in addressing the deficit.

The relationship between studies of normal language processing and aphasia is a symbiotic one. Studies investigating the breakdown of language processes must be grounded in a theory of normal language processing. At the same time, evidence from language disruption, both normal and aphasic, is one of the most important contributors to normal language processing theory. Speech errors produced in both normal and aphasic discourse have provided clues about the underlying mechanisms of lexical retrieval during language production. In addition to informing normal language processes, aphasic error studies have also addressed the goal of describing the mechanisms of language breakdown, in order to learn more about the characteristics of aphasic syndromes. However, much remains unclear about the nature of aphasic speech production deficits, and their implications for models of language production.

For example: What are the specific factors which make particular lexical items vulnerable to error, and what are the factors which determine the ultimate form of the errors produced? How can the study of aphasic speech production inform our understanding of normal speech production? At what level of production might the deficits be located, or would deficits be better characterized as more global processing impairments? How do the patterns of speech errors shown by individual aphasic subjects relate to clinically defined syndromes?

The current study addresses these questions through an examination of the phonological errors produced by an unselected group of aphasic individuals in both spontaneous and structured speech tasks. The form and frequency of occurrence of different types of errors are analyzed within the context of the speech sample, as a means of determining how the linguistic context contributes to error production in running discourse. The corpus of aphasic errors is compared to a corpus of normal speech errors to identify qualitative and quantitative differences. Error distributions in the individual speech samples are also examined for distinctive patterns that might be related to aphasia profiles. This descriptive analysis forms the basis for the focus of the investigation—the role of phonological relationships in lexical access, and how they contribute to error production in aphasia.

The naturalistic and experimental tasks in the current study address a structural aspect of the lexicon which is presently receiving a fair amount of attention in psycholinguistic research with non-brain-damaged subjects, that is, the role of the phonological 'neighbourhood', or the set of phonologically related words with which a target is assumed to compete for lexical selection. Characteristics of the phonological neighbourhood, such as the number of words which are phonologically similar to a target (neighbourhood density), and the frequency of occurrence of those neighbours (neighbourhood frequency) relative to the target's own frequency, are analyzed in spontaneously produced errors, and compared to a corpus of correctly produced targets. The influence of these factors on the accuracy of picture naming is also examined, in order to replicate and extend findings from spontaneous speech. Examining factors that have not been considered before in reference to lexical access deficits of aphasia may help to reveal undiscovered mechanisms underlying some of the more abstruse aphasic errors. In addition, replicating normal speech-error analyses on corpora of aphasic speech errors will provide some insight into normal speech production processes and their vulnerability to breakdown in both non-brain-damaged and aphasic speakers.

Overview

In order to place the study of phonological speech errors into a larger context, the first chapter (**Errors as Evidence**) describes the evolution of normal and aphasic speech error investigations, outlining the methodological difficulties entailed and how they have been addressed. In the second chapter (**Linguistic Constraints on Error Production**), principal findings of error research are reviewed in reference to how they have contributed to our understanding of normal and aphasic speech production processes. The third chapter (**Speech Production Models**) provides a summary of the major models of language production and the main issues that have shaped their development. Particular attention is paid to theories of how lexical items are stored and accessed during language production. The final section in Chapter 3 (Similarity Neighbourhoods) focuses on the role of phonological neighbourhoods in lexical access, first in word recognition research, then in speech production research. These chapters set the stage for the current investigation. In Chapter 4 (**The Pilot Study**), a preliminary study is described, in which existing error data was analyzed in order to direct the methodological procedures of the main study. The methods and results of this principal investigation are presented in Chapter 5 (**The Main Study**). Finally, the sixth chapter (**Neighbourhoods in Aphasic Speech Production**) presents a discussion of the results of the current study with reference to previous findings in the literature. In this chapter, theoretical and clinical implications are also discussed.

Chapter 1. Errors as Evidence

The primary goal of the past century of speech-error research has been to reveal the structures and processes of normal language production, by investigating the factors which appear to promote errors, and those which appear to restrict their occurrence. Following in the footsteps of Hughlings Jackson and others, researchers have made use of "[t]he general strategy...of inferring relevant properties of an unobservable system on the basis of its output characteristics" (Boomer & Laver, 1968, p. 3). Meyer (1992) cautions that error analyses are not sufficient to formulate a comprehensive model of phonological encoding, and advocates a greater reliance on studying error-free speech. However, investigators are in general agreement that the characteristics of errors produced by both normal and aphasic subjects in spontaneous and experimental tasks have provided valuable information to complement findings from studies of normal language production.

This chapter introduces the domain of speech-error research with an overview of the methodology involved in the study of speech errors, the difficulties inherent in error collection and classification, and the steps taken to overcome these difficulties. In discussing issues of classification, a general outline of error typologies which have been described for normal and aphasic speakers is presented. The terminology used in both domains is explained, and examples are given for the various types of errors observed. Thus, this section is intended to provide the reader with a frame of reference for the error research discussed in subsequent chapters.

Error Collection

The method used to gather speech errors has an impact on the representativeness of the errors studied, and the validity of inferences which can be drawn from their occurrence. The contexts in which errors are collected have been recognized to influence both the incidence of error occurrence and the patterns of errors observed (Stemberger, 1992). Such observations are also highly dependent on the way in which errors are classified (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997b), a fact which has, in turn, guided methods of error collection.

Speech errors are collected by an increasing variety of paradigms, ranging from completely unstructured to highly unstructured, each with its attendant advantages and disadvantages (see Cutler, 1981; Meyer, 1992; Stemberger, 1992, for reviews). These can be divided into two general methods—the systematic analysis of spontaneous speech, and the analysis of experimentally elicited responses (Garrett, 1980). The nature of the speech produced in each type of context can vary due to such factors as the subjects from whom the errors are collected, the naturalness of the task, and the vocabulary constraints of the situation (or, in the case of experimental tasks, the characteristics of the stimuli used). Furthermore, the nature of the errors perceived and recorded from the speech produced can be influenced by the method of collection used, as well as intrinsic factors of the speech structures themselves.

Spontaneous vs Elicited Errors

Most early studies of errors produced by non-brain-damaged subjects were qualitative in nature, citing examples overheard in natural discourse (Dell & Reich, 1981; Fay & Cutler, 1977; Fromkin, 1971; Harley, 1984; MacKay, 1970a; Meringer & Mayer, 1895; Nooteboom, 1973; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1985). This method has the advantages of being usually quite unobtrusive (Harley, 1984) and having 'face validity', meaning that the errors collected can be considered representative of real speech errors (MacKay, 1980). Furthermore, all levels of language production (e.g. phonological, syntactic, semantic, and pragmatic) are represented in spontaneous speech (Harley, 1984). Variations on this method include gathering errors from tape-recorded samples of discourse from, for example, conversations (Svartvik & Quirk, 1980), radio discussion panels (Ferber, 1991), conference presentations and psychiatric interviews (Boomer & Laver, 1968). Most corpora, however, have been simply recorded in writing by the experimenter at the time the error is committed or shortly thereafter.

Characterized by Garrett (1980, p. 180) as the 'catch-as-catch-can' technique, this method also has several short-comings. One is that, in most studies (but see Meringer & Mayer, 1895; Harley, 1984, for notable exceptions), there is no permanent record taken of the context of the error, which limits the ability of the experimenter to deduce the true cause of the error (Fay & Cutler, 1977; Harley, 1984; Kohn & Smith, 1990; Kupin, 1982). Determining the nature and source of the error is also limited by the ability of the listener to discern the speaker's intended target. Furthermore, if the experimenter is engaged in interactive discourse with the speaker from whom errors are being collected, the very act of error collection may introduce a 'participant-as-observer' effect into the context (Kupin, 1982). Another disadvantage is that potential sampling biases may be introduced, due to the selective conditions under which errors are collected (Laubstein, 1987; MacKay, 1980). One of the most severe criticisms levelled at spontaneous speech corpora is that they are also highly vulnerable to perceptual biases (e.g., see Mowrey & MacKay, 1990). The latter two points will be taken up in further detail in the following sections.

Despite their drawbacks, such error studies have been instrumental in illustrating the types of errors that occur in natural (or 'naturalistic') speech situations, and have paved the way for more controlled studies. More recently, paradigms designed to elicit certain types of errors have become more widely used (e.g. Baars, 1992a; Baars, Motley, & MacKay, 1975; Dell, 1984; Levelt et al., 1991a; Levitt & Healy, 1985; Martin, Weisberg, & Saffran, 1989; Schriefers, Meyer, & Levelt, 1990; Shattuck-Hufnagel, 1992), in part to compensate for the disadvantages of spontaneous speech studies. One of the most commonly used techniques is the often speeded repetition or oral reading of 'tongue-twisters' made up of either real words (e.g. Shattuck-Hufnagel, 1992) or non-word strings (e.g. Kupin, 1982; Levitt & Healy, 1985; Sevald, Dell, & Cole, 1995; Shattuck-Hufnagel, 1992; Vitevitch, ms in prep). In one variation of this method, called the SLIPs technique (*Spoonerisms of Laboratory Induced Predisposition*, e.g. Baars, 1992a; Baars et al., 1975; Dell, 1984; Motley & Baars, 1975), initial consonant reversals are stimulated by presenting word pairs with the same initial consonants (e.g. *ball doze, bash door, bean deck, bell dark*), which bias the production of a target word pair with the opposite pattern of initial consonants (e.g. *darn bore*). Other techniques involve speeded naming (e.g. Levelt et al., 1991a), the description of an array of items selected for their semantic or phonological confusability (Levelt, 1983; Martin et al., 1989), and naming in a picture-word interference paradigm (Schriefers et al., 1990).

Such structured tasks have allowed experimenters to manipulate certain parameters of the stimuli that spontaneous error studies have shown to be relevant, such as frequency of occurrence, syllabic structure, and grammatical class, in order to test specific hypotheses. In spontaneous speech studies, a large amount of speech must be monitored in order to gather a corpus of errors which is sufficient for analysis. In experimental tasks, however, aspects of the task such as rate of speech and the ability to self-monitor, and aspects of the stimuli such as repeated phonemes within a phrase or list of words, may be manipulated to elicit more errors (e.g. Baars, 1992a; Dell, 1984; Levitt & Healy, 1985; Shattuck-Hufnagel, 1992). In addition, the number of opportunities for certain types of errors to occur may be controlled, allowing more accurate measurement of their relative incidence (Levitt & Healy, 1985).

These advantages are offset by the possibility that errors produced in an experimental situation may be artifacts of the elicitation technique, and thus may not be representative of spontaneously produced errors (Bierwisch, 1981; Dell, 1990; Fromkin, 1980; Garrett, 1976; Levitt & Healy, 1985; MacKay, 1980). In fact, many error elicitation techniques (e.g. Baars, 1992a; Baars et al., 1975; Motley & Baars, 1976a) devise some sort of 'trick' to promote errors, such as diverting the subject's attention from the production task, or creating expectations which are then violated (Kupin, 1982). Meyer (1992) points out that, in experimental tasks, "some of the normal planning processes might be omitted or altered and that the articulation might be more difficult than in spontaneous speech" (p. 197). Garrett (1980) adds that "experimentation inevitably involves the risk of confounding comprehension processes with putative production processes" (p. 178), a caution that may be particularly relevant to the elicitation of errors from aphasic subjects.

Investigations of aphasic errors are prone to the same difficulties as normal error studies, as well as some additional ones. Spontaneous speech tasks have also been used in aphasic error studies (e.g. Blumstein, 1973a); however, aphasic speech sampling is one step removed from 'natural' by virtue of the artificial context in which samples are collected, and the contrived relationship between patient and clinician, or subject and researcher. One advantage that aphasic error studies have over normal error studies is the frequency with which errors occur (e.g. Béland, Caplan, & Nespoulos, 1990; Stemberger, 1982b). Thus, investigators of aphasic speech have the option to use tasks which are more structured than spontaneous speech, but not manipulated specifically to elicit errors, tasks in which normal subjects would be expected to produce very few errors. In tasks such as picture naming, repetition or oral reading, for example (e.g. see Kohn & Smith, 1994a; Kohn & Smith, 1995; Kohn, Smith, & Alexander, 1992), the targets are pre-determined so that the relationship of errors to targets may be analyzed more easily (Dell et al., 1997b), although it is still sometimes difficult to unambiguously identify the subject's *intended* target. Error elicitation techniques have also been used with aphasic subjects (e.g. Dressler, 1979), not so much to induce a greater number of errors, but to investigate the role of specific linguistic factors in the errors induced. As in normal studies, there is the risk that errors produced in laboratory tasks may not be representative of spontaneously produced error production.

To ensure the validity of errors induced in this way, it is widely recommended that experimental findings be confirmed with independent evidence from more natural contexts (Bock & Levelt, 1994; Cutler, 1981; MacKay, 1980; Meyer, 1992; Stemberger, 1985). Some investigators advocate the use of observational evidence as primary data, to be corroborated by experimental findings (e.g. Stemberger, 1985); others prefer to focus on experimental studies, and validate results by comparison with observations from spontaneous speech (e.g. Meyer, 1992). Most agree that the two approaches are "naturally complementary" (Garrett, 1980, p. 178), in that the disadvantages of one are offset by the advantages of the other. Such comparisons that have been done to date have found "broad similarities" in the patterns of errors observed in natural speech and elicited in experiments (Stemberger & Treiman, 1986), although statistical differences have been noted in the distribution of specific types of errors (Stemberger, 1985). In a review of the similarities and differences among experimental and spontaneous speech findings, Stemberger (1992) notes that results from the two paradigms are "remarkably convergent" (p. 210).

In the past decade, a new paradigm for eliciting errors has become increasingly more popular—the simulation of errors by computational connectionist models. Of course, errors produced by a computer are completely artificial, but the ability of computational models to re-create patterns of errors observed in normal and aphasic subjects (Dell et al., 1997b) has provided another source of evidence to support the study of naturally occurring errors. The contributions of such models will be discussed further in Chapter 3.

Sampling Biases

Subject Sampling

Although it is usually assumed that the normal 'subjects' of observational studies represent a random sample of the population (e.g. Blumstein, 1973b), it has been noted that individual speakers vary greatly in their susceptibility to error (Garrett, 1980; Laubstein, 1987). This problem is particularly acute for spontaneous speech studies, in which subjects are 'selected' by their propensity to produce errors, and the coincidence of being in the company of the experimenter at the time. (On the other hand, Meringer and Mayer (1895, cited in MacKay, 1980) considered the collection of errors from selected speakers who were particularly prone to speech errors to be, not a problem, but a convenient strategy to facilitate error collection.) Dell and Reich (1981) pointed out that the majority of corpora, because they are gathered by only one or two investigators, include errors from a restricted sample of the investigator's most common conversational partners. They avoided this source of bias in their own study by using errors collected by about 200 students.

The heterogeneity of patient populations introduces an added obstacle to obtaining a representative corpus of errors in aphasia studies. The types of errors produced by aphasic subjects, and their distributional patterns, vary widely across the aphasic population (Buckingham, 1980). Few aphasia investigators hold any illusions about the random sampling of their aphasic subjects, but it is a limitation of the aphasic literature that most of the data remains "scattered among case studies" (Kohn & Smith, 1994a, p. 75), and that such studies often focus on unusual cases (e.g. Best, 1996; Blanken, 1990). This method of subject selection makes it difficult to generalize findings to the population as a whole and leaves open the possibility that observations are anomalous. Comparing across aphasic error studies is also difficult because they differ in subject selection criteria. Certain types of aphasic subjects, usually those at the extremes of the severity continuum, may be excluded from study because their errors are too few or too many. For example, non-fluent aphasic subjects are often excluded from error studies in order to factor out the potential confound of articulatory deficits (Dell et al., 1997b; Gagnon, Schwartz, Martin, Dell, & Saffran, 1997). In spontaneous speech studies, global and Broca's aphasics are routinely excluded because of the paucity of their expressive output. It has also been shown that the time post-onset of aphasia at which subjects are tested influences the pattern of errors they exhibit (Buckingham, 1987; Butterworth, 1992; Kohn et al., 1992; Kohn, Smith, & Alexander, 1996).

Error Sampling

In addition to the risk of subject sampling bias, there is also the potential for sampling bias in the types and frequencies of the errors produced, especially in spontaneous speech studies. By their nature, spontaneous speech samples do not provide equal opportunities for all types of errors to occur because of the distributional properties inherent in the language, and the situational context of the error collection (Cutler, 1981; Laubstein, 1987; Levitt & Healy, 1985; MacKay, 1980). MacKay (1980) calls this the "fragmentary data problem" (p. 324). Thus, conclusions regarding relative frequencies of errors must take into account the opportunities available for such errors to occur. In addition, conclusions based on null findings must be made cautiously (Cutler, 1981), keeping in mind the possibility that a more extensive sample, or a sample gathered under different conditions, might turn up examples of the error in question. (See, for example, the controversy concerning the existence of phonotactic violations discussed in the next chapter.) To minimize the fragmentary data problem, it is necessary to gather large samples of spontaneous speech (MacKay, 1970a; 1980). As MacKay warns, "The complexity of speech errors shows that a large number of uncontrollable factors can determine any one error, and we now advance hypotheses only when examples greatly outnumber counterexamples" (1980, p. 320).

This cautionary note is also important for the study of aphasic subjects, from whom reliable and unambiguous errors are extremely difficult to obtain, especially in spontaneous speech situations. Concomitant speech and language disorders may render the output difficult to transcribe, let alone analyze, and the context in which the errors of interest occur may be as abstruse as the error elements themselves. In addition to elements that are not produced correctly, there may be elements that are not produced at all, a type of error by omission, which is obviously difficult to interpret (Dell et al., 1997b). Such omissions, and the exclusion of untranscribable sections of speech samples (e.g. Kohn, 1984) reduce the representativeness of the errors that are analyzed. In addition, aphasic error production is notoriously inconsistent, such that repeated testing in a variety of situations is necessary to ensure that the range of errors characteristic of a particular aphasic subject is fully represented (Béland et al., 1990; Butterworth, 1992).

Misperception and Perceptual Bias

In addition to limitations on the speech errors produced, the errors that are *collected* may represent only a subset of the errors *produced*, for a variety of reasons (Browman, 1980; Cohen, 1980; Cole, Jakimik, & Cooper, 1978; Dell & Reich, 1981; Fromkin, 1971; Games & Bond, 1980; Laubstein, 1987; Mowrey & MacKay, 1990; Stemberger, 1992). (See also Cutler, 1981; Ferber, 1991; and Kent, 1996, for comprehensive reviews.) Again, this is a problem that manifests itself most in the collection of spontaneous speech errors. Listeners are not always reliable in their perception of running speech, especially when that speech derails. Errors may be completely missed due to the listener's inattention or to the imperceptibility of the error (Bawden, 1900; cited in MacKay, 1980; Laubstein, 1987). Ferber (1991) provided a striking demonstration of this by comparing the numbers of errors recorded 'on-line' (i.e. while listening to the speech sample) by four listeners, all of whom were familiar with speech-error analysis (including Ferber herself), to those recorded 'off-line' (i.e. while stopping and rewinding a tape of the same speech sample) by Ferber. Only about one-third of the 51 speech errors recorded off-line were noticed on-line. Ferber noted that errors often co-occurred in clusters of two or three, suggesting one reason for the listeners failing to detect so many errors—having to divide one's attention between listening to the speech sample and recording the errors. Furthermore, of those errors detected, only about half were recorded accurately on-line.

Even more disturbing is evidence that certain types of errors are more salient to the listener, resulting in a higher rate of detection and/or greater accuracy in recording. Studies have shown that errors are accurately perceived in stressed syllables more often than in unstressed syllables (Browman, 1980; Cohen, 1980; Games & Bond, 1980); on word-initial segments more often than on word-final or word-medial segments

(Browman, 1980; Cole et al., 1978; Tent & Clark, 1980); and on consonants more often than on vowels (Cohen, 1980). Non-phonemic errors (i.e. semantic and syntactic errors) have been found to be more easily detected than phonemic errors (Browman, 1980; Tent & Clark, 1980); anticipations more detectable than perseverations (Cohen, 1980; Tent & Clark, 1980); and place-of-articulation errors more detectable than voicing errors (Cole et al., 1978). (Ferber (1991) claims that her results do not support the general hypothesis that some errors are more detectable than others. If this were the case, errors would be expected to show a significant discrepancy in their detection rates, but the vast majority (89%) of the errors found on-line were reported by only one or two of the four listeners and none of errors were reported by all four listeners. However, her sample of errors (only 51 in total) is too small to carry much weight in this regard.)

Higher-level biases help to explain the perceptibility (or lack thereof) of some types of errors (Browman, 1980; Garnes & Bond, 1980). Contextual predictability can make an error less detectable. For example, final consonants are more predictable than initial consonants because the phonetic information available in the beginning of the word biases the upcoming consonant (Cole et al., 1978; Dell, Juliano, & Govindjee, 1993); because the phoneme is more predictable, the listener relies less on the actual incoming phonetic information, and errors are more likely to be missed. Garnes and Bond (1980) describe an experiment in which errors spliced into spoken sentences (e.g. *Check the calendar and the bait*) went undetected because of the expectations created by the semantic context. Other phenomena in which high-level expectations influence phonological perceptions include the well-known phoneme-restoration effect (Warren, 1970), lexical biases in phonetic categorization experiments (e.g. Boyczuk & Baum, 1999; Burton, Baum, & Blumstein, 1989; Fox, 1984; Ganong, 1980), as well as puns and the punch-lines of many jokes. Games and Bond (1980) cite a line of Groucho Marx's as an example: "When shooting elephants in Africa, I found the tusks very difficult to remove, but in Alabama, the Tuscaloosa" (p. 236).

One of the most controversial implications of such findings for speech-error research involves the reality of phonotactic constraints. Although it is generally reported that speech errors obey the phoneme sequencing rules of the language in which they occur, it has been suggested by some that listeners may be perceptually biased to overlook phonologically deviant utterances (e.g. Hockett, 1967, cited in Cutler, 1981; Mowrey & MacKay, 1990). Certainly this is often the case for naive listeners in semantically biased situations, as described above, but this claim goes further in stating that even experimenters trained to listen for errors will fail to detect most phonotactic violations. Mowrey and MacKay (1990) took electromyographic (EMG) measurements of their own tongues and lower lips while producing tongue twisters, and compared these to transcriptions of audio recordings of the same tongue twisters. Results indicated abnormalities in articulatory movement even during the production of segments which sounded completely normal to them, and showed that errors were gradational in character rather than all-or-none phenomena. They postulated that listeners "regularize and idealize" actual speech productions, and that even trained listeners are often "unable to mentally reconstruct the actual sound sequence" (p. 1308). Thus, speech anomalies which do not conform to the listener's percepts of 'phonotactic grammaticality' often go undetected.

The difficulty of accurately perceiving phonetic distortions also poses a serious potential problem for the study of aphasic speech errors. According to Buckingham and Yule (1987), "many subphonemic articulatory aberrations produced by aphasic speakers are perceived by hearers as higher level phonemic substitutions—

substitutions quite often never intended by the aphasic" (p. 113), a phenomenon they call 'phonemic false evaluation' or PFE, after Trubetzkoy (1939, cited in Buckingham & Yule, 1987). It has been observed that most phonological errors in aphasia consist of single-phoneme changes, and most of those differ by only one feature (Blumstein, 1973a), so it may be that many of these constitute misperceived phonetic alterations rather than whole phoneme substitutions. However, in one study comparing acoustic analyses of paraphasic errors, self-corrected productions, and initially correct targets produced by a conduction aphasic it was shown that "most perceived substitutions exhibited acoustic characteristics appropriate to the substituted sound, and thus most likely reflect true phoneme selection errors" (Baum & Slatkovsky, 1993, p. 207). Nevertheless, PFE represents an ever-present threat to the validity of phonemic error studies, requiring investigators to be vigilant in their methodological procedures.

Tape-recording of speech samples (e.g. Boomer & Laver, 1968; Svartvik & Quirk, 1980) reduces the chance of mishearing or overlooking errors (Ferber, 1991; Levitt & Healy, 1985), but according to some researchers, the auditory signal is insufficiently reliable, and should be supplemented with acoustic and/or physiological measurements (Kent, 1996; Mowrey & MacKay, 1990). However, as the time, equipment and expertise required to make use of such instrumental analyses are frequently prohibitive, a more common solution has been to ensure a minimum level of reliability of transcription through intra-judge and inter-judge comparisons; that is, by having multiple listeners make multiple 'passes' through the audiotaped sample. In one study, acoustic analyses and auditory perceptual judgements were used to identify phonemic paraphasias (Shinn & Blumstein, 1983). Using a criterion of 100% agreement among four phonetically trained judges to identify errors as 'reliable', the number of perceived phonemic paraphasias was reduced from 300 to only 11; these matched spectral templates for good exemplars of the phoneme in question. One could argue, however, that such a criterion may be too strict. Just as it is not necessary for productions to be 'good exemplars' to be considered productions of the *intended* phoneme, it is also not necessary for productions to be good exemplars to qualify as phoneme substitutions. Although one cannot be 100% sure of the intended phoneme in such cases, a too-strict criterion will under-estimate the incidence of phonemic paraphasias.

The precaution of establishing the reliability of transcriptions is common-place in studies of aphasic errors (e.g. Bastiaanse, Gilbers, & van der Linde, 1994; Canter, Trost, & Burns, 1985; Gagnon et al., 1997; Goodglass et al., 1997; Kohn, Melvold, & Shipper, 1998; Shinn & Blumstein, 1983), but surprisingly rare in normal error studies (but see Boomer & Laver, 1968), except where reliability is the focus of the study (e.g. Ferber, 1991; Mowrey & MacKay, 1990). Admittedly, the threats to reliability are much greater for aphasic error studies, where the frequency of errors and the potential for confounding articulatory distortions necessitates a more careful assessment of the true nature of phonological errors.

Error Classification

Once an error corpus is collected, the first step before being able to make any inferences about the processes of normal language production is to describe the error as fully as possible—the linguistic level involved, the mechanism or process by which the error is produced, and the presumed source of the error (Dell, Burger, & Svec, 1997a; Stemberger, 1985). Only then can conclusions be drawn about how often specific types of errors occur, in what contexts they are most likely to occur, and what factors contribute to their occurrence (to be discussed further in the following chapter). Many of the types of errors observed in normal speakers are also produced by aphasic speakers, and the principles of classification are thus relevant to both populations. But there are also some patterns of error production which are characteristic of specific clinical sub-types of aphasia. These will be reviewed briefly, following a description of normal speech errors.

Normal Speech Errors

Baars defines normal slips of the tongue as "errors that violate their own governing intentions" (1992b, p. vii; see also Boomer & Laver, 1968). In other words, the correct intentions formulated at one stage during the production of a utterance somehow get derailed in the process of being passed on to the next stage. Although errors can occur at all levels of language production, from the pragmatic intentions of the message to the articulatory movements required to produce the utterance, linguistic analyses generally restrict their focus to the stages of sentence formulation, lexical selection and phonemic encoding, involving units as large as phrases down to phonemes and phonetic features. Examples of errors at each of these levels are presented in Table 1-i (following page). (There is also evidence that errors may occur at the level of stress assignment (e.g. Cutler, 1980; Fromkin, 1971), although these types of errors will not be discussed here.)

Errors are further described in terms of their relationship to the target. At syllable and segment levels, errors are by definition phonologically related to the target, because only a portion of the target is produced in error. Specific types of phonological relationship are differentiated by the mechanisms giving rise to the errors, and the degree of overlap between the target and the error. At the lexical and phrase levels, errors may be phonologically related to their targets (see examples 1a, 2a, b, d, and e in Table 1-i), or there may be a semantic relationship (example 2c), a syntactic relationship (examples 1b and 3), or no apparent relationship at all (example 2f).

Table 1-i. Linguistic Levels of Error Occurrence

1. Phrase Level

   a) *what came over/took hold of me > what took over, overtook me**

   b) *a far better man than anyone here > a farther man than anyone better here***

2. Word Level

   a) *he chooses/takes > he chocks**

   b) *she'd burnt > she'd burst**

   c) *last year > next year**

   d) *Get out of the car > Get out of the Clark****

   e) *I haven't a clue > I haven't a cue****

   f) *I've read all my library books > I've eaten all my library books****

3. Morpheme Level

   a) *historical interest > historical interested**

   b) *transcriptions > transcript_s**

   c) *in conclusion > in concludement***

4. Syllable Level

   a) *a degree > a _gree**

   b) *pussy cat > cassy put***

   c) *foolish argument > farlish***

   d) *butterfly and caterpillar > butterpillar and caterfly***

5. Phoneme Level

   a) *thunderous applause > thunderous apprause**

   b) *much more > mich more**

   c) *drugs > d_ugs***

   d) *play the victor > flay the pictor***

   e) *an eating marathon > a meeting _arathon***

6. Feature Level

   a) *define > devine***

   b) *clear blue sky > glear plue sky***

   c) *tab stops > tap stobs***

(from: *Garnham, Shillcock, Brown, Mill, & Cutler, 1981, **Fromkin, 1971; and ***Harley, 1984)

One type of word-level error that has received a fair amount of attention (e.g. Fay & Cutler, 1977; Zwicky, 1982) consists of a real-word substitution which is phonologically but not semantically related to the target, called a 'malapropism' (example 2b). As Zwicky (1982) notes, malapropisms which result from slips of the tongue should be distinguished from 'classical malapropisms', that is lexical substitutions which, although incorrect, are nevertheless *intended* by the speaker (as exemplified by the original Mrs. Malaprop invented by Sheridan, 1906). Because they do not reflect disruptions in *on-line* phonological processing, classical malapropisms will not be considered here. (Although it is recognized that such errors may well exist in corpora of natural speech errors, particularly those collected anecdotally, such instances are probably relatively rare.)

Thus, the error/target relationship is an indication of the level at which production has derailed, although the decision as to which level an error should be assigned is often ambiguous. For example, Fromkin (1971) acknowledges that many of the feature changes observed in her database might be classified instead as phoneme changes, because they always result in a different but existing phoneme. If that were the case, however, it would be necessary to explain an apparent feature reversal such as 6b as two separate phoneme substitutions, one involving a change from a voiceless to a voiced counterpart, the other involving the complementary change from voiced to voiceless counterpart (Fromkin, 1971). On the other hand, the relatively low incidence of featural errors has been cited as evidence for the indivisibility of segments (Shattuck-Hufnagel, 1979).

Similarly, phonologically related word substitutions, or malapropisms, such as 2b may also be classified as phoneme substitutions. In some studies, this sub-lexical explanation seems intuitively more likely; for example, the SLIPs technique (Baars et al., 1975; Motley & Baars, 1975) is specifically designed to elicit sub-lexical errors which may, by choosing appropriate stimuli, create real words (e.g. *darn bore* produced as *barn door*). On the other hand, given equal opportunities for the production of real-word and non-word spoonerisms, real-word outcomes are more likely (Baars et al., 1975), suggesting that they are true lexical substitutions. Where the creation of equal opportunities is not feasible (as in analyses of spontaneously produced errors), investigators compare the incidence of real-word over non-word production in the experimental corpus to a 'pseudo-corpus' created to estimate chance probabilities of lexical and non-lexical outcomes (e.g. Dell & Reich, 1981; Dell et al., 1997b; Martin, Gagnon, Schwartz, Dell, & Saffran, 1996; Stemberger, 1985). Some suggest, however, that this lexical bias reflects the operation of a *post-hoc* filtering function performed by an output monitor or editor (e.g. Baars et al., 1975; Buckingham, 1980; Garnsey & Dell, 1984; Levelt, 1989; Levelt et al., 1991b), in which case a phoneme-level explanation of the error would still be workable.

These issues will be discussed in more detail in the next section, but a general rule of thumb in determining the unit involved in the error is to assume that the simplest possible mechanism is at work (Stemberger, 1985). In Fromkin's (1971) example described above, a feature reversal is a more parsimonious explanation than postulating two independent phoneme substitutions. Furthermore, the incidence of such errors relative to other types of errors and to chance expectations provides information that helps to disambiguate the level involved (Stemberger, 1985).

As noted earlier, how production might derail is also an important consideration of error classification. At each of the different levels at which errors may occur (see Table 1-i), linguistic components might be substituted (e.g. 2c, 3c, 4d, 5a, 6a), added (e.g. 3a), deleted (e.g. 3b, 4a, 5c), or blended (e.g. 1a, 2a, 4c). Substitutions occur when a word or phoneme is misselected from the lexical store or from elsewhere in the utterance under construction; additions and deletions can be considered substitutions involving null elements (e.g. Dell, 1986; Stemberger & Treiman, 1986). Blends occur when parts of two words competing for selection are simultaneously produced (but see Laubstein, 1987, 1999, for an alternate explanation). Among phoneme errors, substitutions have been found to be the most common type of error for both normal (e.g. Boomer & Laver, 1968; Garnham et al., 1981; Shattuck-Hufnagel & Klatt, 1979) and aphasic subjects (e.g. Blumstein, 1973a; Buckingham, 1977; Burns & Canter, 1977; Christman, 1994; Green, 1969; Martin, Wasserman, Gilden, Gertman, & West, 1975; Miller & Ellis, 1987; Romani & Calabrese, 1998; Trost & Canter, 1974).

For errors with an identifiable source within the phonological context of the utterance, the directionality of the influence is an informative aspect. Substitutions and additions can be broken down into anticipations (e.g. 4c), perseverations (e.g. 5a), or shifts (e.g. 5e). Strictly speaking, a shift occurs when a segment is *moved* out of its original spot and into another, whereas anticipations and perseverations occur when segments remain in their intended position, but are also *copied* into a new position. In practice, however, the two types of errors are often confused, especially in self-interrupted utterances containing an anticipatory shift, where it is unclear whether the complete utterance would have contained one or two instances of the anticipated phoneme. Exchanges (e.g. 4b, 6b), also called transpositions, metatheses, or spoonerisms, can be explained as either a right-to-left (i.e. anticipatory) shift and a left-to-right shift to fill the gap, or as two separate substitutions. The former mechanism, which suggests a dependence between the two operations, is a more parsimonious explanation.

Such contextually influenced (or 'movement') errors involve the misordering of elements, while non-contextual (or 'no-source') errors involve the misselection of linguistic elements (Bierwisch, 1981; Dell et al., 1997a). Contextual and non-contextual errors are also referred to as 'syntagmatic' and 'paradigmatic' errors, respectively (e.g. Dell, 1986; Talo, 1980), particularly in the aphasic literature (e.g. Buckingham, 1986; Lecours & Lhermitte, 1969) after Jakobson's dichotomy of language functions (Jakobson, 1956). Although it has proven a useful distinction in error studies, Dell and colleagues (1993) point out that "[t]he actual breakdown between movement and nonmovement slips...depends heavily on how error sources are defined. ...the distinction is not clearcut, and hence, we find it profitable to view a contextual influence as graded" (p. 184). Furthermore, it is often unclear whether or not an error is contextually determined, in part because it is not known over what distance contextual elements of various sizes can exert an influence, nor whether context exerts similar effects on different types of errors (Schwartz, Saffran, Bloch, & Dell, 1994). Nevertheless, reference to as much of the context as is available "often permits disambiguation between alternative interpretations" (Harley, 1984, p. 195; see also Fay & Cutler, 1977; Kohn & Smith, 1990).

For higher-level errors, that is, errors at a conceptual level of planning, a similar distinction has been made between errors whose source can be traced to the planned utterance, called 'plan-internal' errors, and those for which no source within the planned utterance is evident (Meringer & Mayer, 1895, cited in Butterworth, 1981; Garrett, 1980; Harley, 1984; 1990). Such 'non-plan-internal' errors may be due, for example, to: environmental intrusions, as in 2d (Table 1-i), which occurred when the speaker was looking at a shop sign that read 'Clark's'; previous utterances, as in 2e, spoken by someone who had just been discussing how to hold a snooker cue; or unrelated thoughts as in 2f, spoken by someone who was hungry at the time (Harley, 1984). Another possible source of such conceptual intrusions comes from repressed thoughts, giving rise to the 'Freudian slip' (1901, translation 1965, cited in Motley & Baars, 1976a). More recent investigations, however, do not give much credence to Freud's theory (Boomer & Laver, 1968; Ellis, 1980; Motley & Baars, 1976a). These types of errors are often excluded from studies (Butterworth, 1981) or are classified as unrelated, probably because the only way to discern the source of the error is to have access to the full situational context and to the speaker's intuitions about the error.

Aphasic Speech Errors

With an evocative metaphor, Garrett (1992) describes the efforts of aphasiologists to classify aphasic speech errors: "The impulse to tame the polymorphous bestiary of anomias, alexias, paralexias, paraphasias, dysphasias, dyslexias, dysgraphias, and neologisms has occupied many" (p. 143). (One is tempted to divide errors by species, genus and phylum.) Many of the errors observed in aphasic patients correspond to normal patterns, although in some cases different terminology is used. As in normal error corpora, both semantically and phonologically related substitutions occur, and phonological errors may result in either real-word or non-word errors. Real-word substitutions are sometimes called 'verbal paraphasias', or semantic paraphasias if the error and target share some feature of their meaning. Phonologically related real-word substitutions, corresponding to normal malapropisms, are usually designated 'formal paraphasias', or sometimes as 'phonic' verbal paraphasias (Green, 1969). Due to the difficulty of establishing such verbal paraphasias as true word-selection errors, Butterworth (1979) called them 'jargon homophones'. Non-word errors are called 'neologisms', although sometimes this term is reserved for those non-word errors which are unrelated to any identifiable target, while phonologically related non-word substitutions are called 'phonemic' or 'literal' paraphasias. Where the term 'neologism' is used for both types of non-word error, the distinction is made by describing the neologism as either 'target-related' or 'abstruse' (Buckingham, 1990).

The definition of phonological relatedness can vary widely across studies. Some investigators use rather strict criteria, such as that the target and the error must share at least 50% of their component phonemes (Christman, 1994; Mitchum, Ritgert, Sandson, & Berndt, 1990; Nickels & Howard, 1995). According to Nickels and Howard, this criterion "satisfied the intuitive feeling of relatedness in the vast majority of cases" (1995, p. 220). Others require only one phoneme overlap (Best, 1995; Best, 1996), sometimes with the stipulation that it occur in the same word and/or syllable position as the target (Best, 1995; Best, 1996; Gagnon et al., 1997; Harley, 1984; 1990; Kohn, Melvold, & Smith, 1995). Overlap in number of syllables and stress pattern are also used as criteria (Christman, 1994; Harley, 1984, 1990; Kohn et al., 1998). The criteria used carry implications for the classification of phonemic paraphasias, in particular for the distinction between target-related and non-targeted-related neologisms.

Although non-word errors are not abnormal in themselves, the frequency with which they occur in some types of aphasia, and the degree to which they deviate from the target distinguish them from normal errors (Buckingham, 1980; Dell et al., 1997b; Talo, 1980). When a particular non-word utterance is perseverated repeatedly, or when a word or phrase is used repeatedly in inappropriate contexts, it is termed a 'stereotypy' or 'automatism' (e.g. Blanken, Dittman, Haas, & Wallesch, 1988) or a 'recurrent utterance' (e.g. Code, 1982). Although the origin of stereotypical utterances remains a mystery, there are some common clinical speculations; for example, that stereotypies carry particular emotional weight or personal relevance, or that they recur because they were the patient's first post-stroke utterance (Code, 1982). Lecours and

Rouillon (1976) reported a 'predilection' for stereotypies related to work or health among male Wernicke's aphasics, and to family or religion among female Wernicke's aphasics. Code (1982) also noted a number of expletives in his corpus of recurrent utterances from 75 aphasic subjects, particularly from the male subjects.

Discourse that consists almost entirely of non-words or inappropriately used words is called 'jargon', and is characteristic of Wernicke's aphasia. If speech output consists mostly of non-words, it may be called 'neologistic jargon' or 'glossolalia'. Kertesz and Benson (1970; citing Alajouanine, 1956) described three types of jargon: 'undifferentiated jargon', consisting mostly of stereotypies; 'asemantic jargon', consisting mostly of neologisms and empty words, but with a discernible syntactic structure; and 'paraphasic jargon' consisting mostly of real, but semantically inappropriate words. Lecours (1982) observed the recurrent use of 'prediiection segments', or strings of syllables, and a concomitant reduction in the variety of phonemes used, in the glossolalic speech samples of Wernicke's aphasics. These repetitive, perseveratory patterns have also been noted to give the speech of jargon aphasics a certain alliterative and assonantal quality (Buckingham & Kertesz, 1974; Buckingham, Avakian-Whitaker, & Whitaker, 1978; Green, 1969).

Abnormal patterns of repetition also play a role in a pattern of speech errors known as '*conduites d'approches*' or 'sequences of phonemic approximation' (SPAs) (Buckingham, 1992; Joanette, Keller, & Lecours, 1980; Kohn, 1984, 1989; Valdois, Joanette, & Nespoulos, 1989), although here the repetitive attempts are more purposeful and directed. These errors involve repeated attempts at a target, resulting in strings of phonologically related words, non-words, and fragments which tend to show a general progression toward the target (Joanette et al., 1980). Although characteristic of conduction aphasia, they also occur in other types of aphasia. In a comparison of successive approximation errors produced by Broca's, Wernicke's and conduction aphasics, Kohn (1984) found that, although there were no significant group differences in the number of phonological paraphasias produced, or in the number of multiple attempts overall, it was the greater number of attempts at the same target, and the greater number of word fragments, that set conduction aphasics apart from the other two groups. In this study, the conduction aphasics were not found to be more successful than the other groups in the proportion of attempts that eventually achieved target form, although Joanette and colleagues (1980), measuring the degree of success by the phonological distance of the final attempt from the target, found that conduction aphasics did achieve a higher success rate than either Wernicke's or Broca's aphasics.

Another pattern of error not usually observed in normal subjects is the dysarthric (and/or apraxic) distortion of phonemes that has been observed to distinguish non-fluent and fluent forms of aphasia (Baum, Blumstein, Naeser, & Palumbo, 1990; Blumstein, 1973b; Blumstein, Cooper, Zurif, & Caramazza, 1977; Tuller, 1984). As disruptions of phonetic implementation rather than phonological encoding, such productions are usually excluded from studies of aphasic speech errors (e.g. Blumstein, 1973b), and will not be discussed here, except in reference to the difficulties involved in distinguishing the two types of error.

One of the main goals of aphasic error studies has been the search for distinctive phonological deficits. While patterns of errors such as neologistic jargon, *conduites d'approches*, and motor speech impairment exemplify broad differences among clinically defined aphasic syndromes, the analysis of paraphasic errors has not been shown to discriminate well among clinical syndromes (e.g. Blumstein, 1973a; Goodglass, Quadfasel, & Timberlake, 1964; Hofmann, 1980; Kerschensteiner, Poeck,

& Brunner, 1972; Kohn, 1984; Mitchum et al., 1990). In her landmark study of spontaneous speech errors, Blumstein (1973b) found no differences between Broca's, Wernicke's and conduction aphasics in the rank order of types of phonological speech errors, or in the degree to which markedness and phonetic distance contributed to the phonemic substitution errors produced. Phonemic paraphasias, in particular, are common across aphasic sub-types (e.g. Blumstein, 1973a), and most typical of perisylvian syndromes (Ardila & Rosselli, 1993; Mitchum et al., 1990; but see Barton, Maruszewski, & Urrea, 1969; Moerman, Corluy, & Meersman, 1983).

However, some group differences have been noted. Anomic aphasics tend to show fewer neologistic errors (Kohn & Goodglass, 1985) than do other aphasics, whereas Wernicke's aphasics tend to show more (Ardila & Rosselli, 1993; Mitchum et al., 1990). Broca's aphasics have been found to be more likely to make errors on initial than final phonemes (Trost & Canter, 1974), whereas fluent aphasics have shown the opposite pattern (Burns & Canter, 1977). In a comparison of these two studies (Canter et al., 1985), Broca's aphasics showed a greater percentage of one-feature changes than the fluent groups (see also Nespoulous, Joanette, Beland, Caplan, & Lecours, 1984). It has been suggested that such group differences noted in some studies may be attributable to the different tasks used (Burns & Canter, 1977), to the feature system used to classify errors (Buckingham, 1987), to severity differences among aphasic groups (Laine, Kujala, Niemi, & Uusipaikka, 1992; Moerman et al., 1983), and to the heterogeneity of subjects within the groups (Laine et al., 1992; Mitchum et al., 1990). In addition, it is apparent that many of the findings of group differences can be attributed to the contribution of articulatory deficits to the patterns of production errors in Broca's and other non-fluent aphasias (Blumstein, 1973a; Blumstein, Cooper, Goodglass, Statlender, & Gottlieb, 1980; Canter et al., 1985; Lecours & Caplan, 1975;

Monoi, Fukusako, Itoh, & Sasanuma, 1983; Tuller, 1984). Another serious limitation on the validity of comparing such results across studies lies in the different systems of classification used to group errors. In more recent studies, following cognitive neuropsychological trends, the focus has switched from the description of clinical syndromes to the description of deficits in terms of levels of disruption in normal production models (e.g. Kohn & Smith, 1994a; Kohn et al., 1996; Laine et al., 1992). These will be discussed more fully in the chapter on speech production models.

Chapter Summary

When speech errors first became recognized as a valuable source of information about language production, studies were largely anecdotal, describing the hypothesized mechanisms underlying a few examples of different types of errors. As the complexity of these mechanisms became apparent, speech error researchers became more systematic in their approaches. In general, methods of collection have evolved from more observational to more experimental, and have been extended from normal to brain-damaged populations. Observations from and experiments with aphasic subjects have proven particularly informative for modeling language production, although the paradigms used in aphasic error studies tend to lag one step behind studies of normal errors.

Awareness of threats to the reliability and validity of errors collected from both spontaneous speech and experimental tasks has vastly improved the study of speech errors in normal and aphasic subjects. Experimental and technological methods of establishing reliability have minimized the impact of sampling and perceptual biases. In addition, the use of converging sources of evidence, from both normal and aphasic speech error studies, using spontaneous speech corpora as well as experimentally elicited and computationally simulated errors, has added credibility to the growing body of knowledge concerning the occurrence of speech errors. The challenge in the study of both normal and aphasic errors is to integrate results from experimental studies with a representative array of observations from more natural speech contexts and from a variety of types of aphasia.

Classification efforts, while still far from achieving any kind of consensus, have become increasingly specific within each domain of study. Broad descriptive categories such as 'semantic errors' and 'phonological errors' have been sub-divided into categories related to the mechanisms hypothesized to give rise to the errors (specifying, for example, contextually influenced phonological errors as anticipatory or perseveratory), and to the nature of the resulting errors (such as word or non-word errors). These methodological advances have occurred, in part, in response to criticisms of earlier less structured approaches, but they have introduced their own methodological complications. "To a considerable extent, error categorization is a theory-laden decision, both with respect to the size of the disrupted unit and the nature of the disruption" (Dell et al., 1997a, p. 124). Thus, the reliability and validity of classification efforts improve as we gain a deeper understanding of the factors which promote and constrain error production. In the next chapter, these factors are reviewed in detail.

Chapter 2. Linguistic Constraints on Error Production

As important as observing what is disrupted in error productions, is observing what is preserved of the intended utterance. Although the levels and mechanisms of speech disruption described in the previous chapter suggest an almost unlimited range of potential errors, there are restrictions on the types of errors that are actually observed, and on the frequency with which certain types of errors occur. Garrett notes that errors bear "a principled relation to production" (1980, p. 217); it is these regularities which provide evidence of linguistic rules operating at different levels of language production. Aphasic error patterns are compared to normal error patterns in order to illustrate not only what has gone wrong in different types of aphasia, but also what is still 'right' with them (Buckingham, 1980). Because they come from a language-disordered population, aphasic errors are both more plentiful than normal errors and, arguably, a more stringent test of hypotheses regarding error production and, by extension, normal language production. As asserted by Boomer and Laver (1968):

> To the degree that observed tongue slips can be shown to be structured, and not simply the result of random malfunctioning of the speech producing process, then their obedience to the constraints of a descriptive and explanatory theory may provide the basis for deriving some of the relevant properties or characteristics of the sequencing system, of interest to linguistics, psychology and neurophysiology. (p. 3)

This chapter presents a review of findings from studies of both normal and aphasic errors which have addressed such empirical questions as: Which types of linguistic structures are vulnerable to error, and which, if any, are invulnerable? For those aspects of speech output which can be disrupted, is the change predictable? What are the constraints on the ultimate form of the error? Are errors subject to linguistic rules, or only to statistical probabilities? To answer such questions, it is essential to distinguish the factors which determine a target's susceptibility to error, or 'slipability' (Dell, 1990) from the factors which determine the nature of the error that occurs (Baars, 1980; Harley, 1984; Kupin, 1982). Fewer studies have focused on the former, in part because fewer investigators have systematically compared error productions to non-error productions, or measured actual error production against the opportunities available for such an error to occur. It is also informative to consider factors related to intrinsic characteristics of linguistic representations separately from factors related to the contexts in which errors occur, although, as mentioned in the last chapter, this distinction is not always clear-cut (Dell et al., 1993). In this chapter, potential linguistic constraints will be described with reference to the linguistic level at which the constraint is presumed to operate, the most important being the levels of lexical, syllabic, and phonological representation. For the moment, the discussion proceeds without reference to a specific theoretical perspective; implications for theories of language production will be addressed in the subsequent chapter on speech production models.

Lexical Factors

It has been hypothesized that characteristics of stored lexical representations and the organization of the mental lexicon have an impact on the ease or automaticity with which words are produced, and thus influence error rates. Such lexical characteristics include semantic factors such as familiarity, imageability and concreteness (e.g. Blanken, 1990; Goodglass, Hyde, & Blumstein, 1969; Kay & Ellis, 1987; Laine et al., 1992; Nickels & Howard, 1995); syntactic factors such as grammatical class (e.g. Dell, 1990; Kohn & Smith, 1993; Zingeser & Berndt, 1990) and morphological structure (e.g. Fromkin, 1971; Goldberg & Obler, 1997); and structural factors such as length and stress pattern (e.g. Best, 1995; Boomer & Laver, 1968; Caplan, 1987; Dell, 1990). (Because the focus of this thesis is on phonological errors, semantic and grammatical influences will not be discussed, except as they relate to processing at a phonological level.) Lexical status itself also turns out to be an important variable in constraining error occurrence.

Lexical Frequency

One of the most robust and long-standing findings in lexical access research has been the influence of lexical frequency on both input and output processes (e.g. Howes, 1954; Oldfield & Wingfield, 1965; Soloman & Postman, 1952). Findings from single-word production tasks show that the effect originates at a lexical level, rather than at the level of articulatory fluency (Jescheniak & Levelt, 1994; Savage, Bradley, & Forster, 1990), or at the level of concept identification (Huttenlocher & Kubicek, 1983; Jescheniak & Levelt, 1994). Furthermore, there is evidence that frequency is encoded at the level of the phonological lexicon (also known as the lexeme level), where phonological word forms are stored, rather than (or possibly in addition to) the semantic lexicon (a.k.a. the lemma level), where word meanings are stored. In picture-naming experiments, Jescheniak and Levelt (1994) illustrated the 'homophone effect': "Low-frequent homophones behaved like high-frequent controls, inheriting the accessing speed of their high-frequent homophone twins. Because homophones share the lexeme, not the lemma, this suggests a lexeme-level origin of the robust effect" (p. 824). This conclusion has also been supported by the finding that a frequency effect was evident for blends only if the blended words were phonologically related (Laubstein, 1999). Similarly, Hotopf (1980) demonstrated frequency effects for form-based substitutions, but not meaning-based substitutions.

Frequent words are produced accurately more often than infrequent words for normal subjects (e.g. Dell, 1990; Harley & MacAndrew, 1992; Stemberger, 1984a; 1986) as well as aphasic subjects (e.g. Blanken, 1990; Ellis, Miller, & Sin, 1983;

Favreau, Nespoulos, & Lecours, 1990; Martin & Saffran, 1992; Pate, Saffran, & Martin, 1987; Romani & Calabrese, 1998; Williams & Canter, 1982). Frequency effects have been shown to affect a variety of error types: spoonerisms from both spontaneous and elicited contexts (Vitevitch, ms in prep); remote and target-related neologisms (Gagnon & Schwartz, 1996); and the production of jargon (Ellis et al., 1983); as well as the occurrence of tip-of-the-tongue states (Harley & Bown, 1998). Some conflicting results exist in the aphasia literature: Blanken (1990) found a frequency effect for naming accuracy overall, but not for a small corpus (n=64) of formal paraphasias produced by a single subject; Martin (1989) found no frequency effect for errors produced in a coloured-shape naming paradigm, but the set of stimuli consisted of only sixteen words; a conduction aphasic studied by Best (1996) showed no frequency effect in picture naming. These results might be explained by the restricted sets of data examined, or by the anomalous nature of the particular cases under investigation. On the whole, however, frequency of occurrence seems to exert a strong effect on the susceptibility of words to error.

However, while influencing an item's susceptibility to error, word frequency does not seem to be a significant factor in determining the outcome of the error, at least not in normal speech (Dell, 1990; Dell & Reich, 1981; Garrett, 1976; Harley & Bown, 1998; Harley & MacAndrew, 1992). There are reports of conflicting findings, however. A recent study by Vitevitch (1997) showed that malapropisms produced in normal spontaneous speech (from Fay & Cutler, 1977) were more likely to have a higher than lower frequency count relative to their targets. On the other hand, Laubstein (1999) found a frequency effect for phonologically related blends, but in the opposite direction—intrusions were more often of lower frequency than their targets, a finding that may be related to the difficulty of determining which word is the target and which is the intruder in blends.

In aphasia studies, results are also equivocal, but suggest a greater impact of frequency on outcome. In a picture-naming study, Gagnon and colleagues (1997) compared the frequencies of occurrence of formal paraphasic responses produced by nine fluent aphasic subjects to the frequencies of a comparable corpus of control words representing chance word-error outcomes. The corpus of formal paraphasias was significantly higher in frequency than the control corpus. Blanken (1990) found no frequency effect on target susceptibility, but also showed a frequency effect on outcome for formal paraphasias produced by one aphasic subject. In another case study, Martin et al. (1994) showed that the real-word errors of a Wernicke's aphasic were higher in frequency than their targets. In a different type of analysis, Code (1982) showed that aphasic errors represent a generally higher-frequency subset of words than normal speech. Over 80% of the real-word recurrent utterances collected in his study occurred at a rate of more than 100 times per million in normal language. Although the comparison has not been made explicitly, it may be that frequency of occurrence has a stronger influence on error outcome for aphasic subjects than for normal subjects.

These results support an early finding by Howes (1964) that the frequency distribution of words in aphasic speech samples, while similar in shape to normal distributions, was shifted towards the higher-frequency end of the spectrum. Unlike the studies described above, which focused on the frequency characteristics of single words, Howes' study looked at spontaneous speech samples. It is a well-known clinical observation that, in connected speech, fluent aphasic patients tend to use many more high-frequency words, such as function words and empty content words, than non-fluent aphasic patients (Goodglass et al., 1969). Furthermore, it has been suggested that contextual probability (or transitional probability), which estimates the likelihood of a word occurring in a given context using a cloze procedure, exerts an influence beyond that of word frequency, and may be a more appropriate type of measure for the connected speech of normal (Beattie & Butterworth, 1979) and aphasic speakers (Goodglass et al., 1969).

Grammatical Class

Syntactic word class also appears to exert an effect on the slipability of words. (Please note that these issues will be afforded greater attention in the discussion of speech production models, but they will be mentioned briefly here.) Content words have been found to be more error-prone than function words for normal subjects (e.g. Garrett, 1975; 1980) and aphasic subjects (e.g. Buckingham & Kertesz, 1974; Butterworth, 1979). Contrary to these results, Kohn and Smith (1993) found no influence of word class on the production accuracy of four fluent aphasics, but proposed that this was due to the single-word production tasks used, which present highly atypical contexts for the production of function words (see also Garrett, 1992). Preliminary data from functors embedded in phrases supported this hypothesis (Kohn & Smith, 1993). Furthermore, in a previous study using a sentence repetition task, a conduction aphasic subject was found to make many more errors on content than function words (Kohn & Smith, 1990).

The grammatical class effect has been attributed to a difference in retrieval mechanisms. It is hypothesized that, whereas function words are retrieved as part of the framework of a sentence during the production of connected speech, content words are selected and inserted into the frame at a subsequent stage, which explains their differential involvement in whole-word errors (Garrett, 1975, 1980, 1992). Furthermore, Dell (1990) appeals to the 'closed-class principle' to explain the relative immunity of function words to segmental error involvement: as members of a closed class (that is, one which does not allow the creation of new units) function words are retrieved as whole units, whereas content words are generated on-line through the operation of linguistic rules and processes relevant to their context.

Despite the intuitive logic of these explanations, it has been alternatively suggested that grammatical class effects are actually artifacts of other differences between the two word classes, such as lexical frequency (Dell, 1990; Ellis et al., 1983; Stemberger, 1984a). Dell (1990) found no differences in error production between content and function words that were homophones, even though they were significantly different in frequency (recall the homophone effect described above, whereby phonological frequency counts for homophones are equivalent). Ellis and colleagues (1983) found no difference between content and function words matched for frequency in oral reading errors produced by a Wernicke's aphasic subject. Furthermore, the grammatical class distinction between content and function words is confounded by differences in phonological structure, such as stress and phonemic content, which have also been shown to influence error production (Buckingham, 1980; Dell, 1990). Whether it is these structural factors or the word class itself which influences their susceptibility to error remains open to debate.

Where grammatical class has a less controversial effect is on the outcome of the error. Many studies have shown that word substitution errors almost always belong to the same syntactic class as their targets (e.g. Bierwisch, 1981; Blanken, 1990; Fay & Cutler, 1977; Fromkin, 1971; Gagnon et al., 1997; Garrett, 1980; Nooteboom, 1973), an effect which falls out of the frame-filler mechanism of sentence construction (Garrett, 1992; Garrett, 1980; Shattuck, 1975). Interacting words in blends also obey the class constraint (Laubstein, 1999). However, violations of this constraint do occur in word misordering errors, e.g. *toast the burn*, (Dell & Reich, 1981); *get a cash checked*, *sudden stops* > *sudden quicks* (Garrett, 1980), but these are attributed to a different level of processing (Dell & Reich, 1981; Garrett, 1980). Even in these errors, though, the distinction between the major classes of content and function words is preserved.

Note that it is always important to consider the relevant domain of the error when defining their constraints (Garrett, 1980). Although word-level errors rarely occur between words of different grammatical class, segmental errors may, and frequently do because of their proximity, e.g. *thunderous apprause** (Garnham et al., 1981); *Bill snovels* show* (Garrett, 1980); *Dan hates milk* > *Dan han* milk*, (Kohn & Smith, 1990). Lexical-level errors are subject to lexical-level constraints, and segmental errors are subject to constraints at the level of phonological structure.

Morphological Composition

Although issues of morphological composition and decomposition are too complex to be discussed in any detail here, there are some relevant results from error research concerning the representation of morphological structure in the lexicon. Findings that illustrate that stems and affixes can act independently in error production (e.g. in the addition of a plural marker to an adjective in the error *sudden quicks* cited above (Garrett, 1980)) provide support for the hypothesis that they have separate representations in the lexicon (Bierwisch, 1981; Butterworth, 1979; Dell & Reich, 1981; Fromkin, 1971; Garrett, 1980; 1988). To use Garrett's (1980) term, affixes can be 'stranded' (e.g. *get a cash checked*). Further support comes from the observation that affixes appear to be correctly applied to neologistic jargon errors in aphasia (Buckingham & Kertesz, 1974; Ellis et al., 1983; Goldberg & Obler, 1997). On the other hand, Stemberger and MacWhinney (1986) found that the frequency counts of inflected verb forms (particularly irregular verbs) affected their error rates, suggesting that at least some words are stored in the lexicon in their inflected form. Morphological inflection also adds an element of complexity to speech production which, at least in aphasic subjects, appears to promote the production of errors, even when phonological complexity is controlled (Martin et al., 1975; Niemi, Koivuselka-Sallinen, & Laine, 1987).

Word Shape

Structural aspects of lexical items, such as their length and stress pattern, also appear to influence error production. (Other phonological factors operating at sub-lexical levels will be dealt with in subsequent sections.) The length of the targeted word has been found to influence the susceptibility of words to error production for normal subjects (e.g. Fromkin, 1971), but particularly for aphasic subjects (Best, 1996; Favreau et al., 1990; Friedman & Kohn, 1990; Howard & Orchard-Lisle, 1984; Kohn & Smith, 1994a; Nickels & Howard, 1995; Pate et al., 1987; Romani & Calabrese, 1998). This finding makes intuitive sense, since the opportunity for error increases with longer words, and longer words also tend to be less frequent (Pate et al., 1987). However, the effect of target length is not so simple. Pate and colleagues (1987) found that longer words were produced less accurately than shorter words by a conduction aphasic patient on an oral reading task, and that this effect was maintained when accuracy was computed as a proportion of the number of syllables produced, rather than a proportion of the number of words produced. Furthermore, strings of monosyllabic words were produced more accurately than multisyllabic words with the same total number of syllables. Thus, the length effect cannot be solely attributed to an increased opportunity for error. Differences in frequency of occurrence, although a contributing factor, could not fully account for the length effect either.

Conversely, Best (1995) reported an aphasic subject who showed a reverse length effect in naming, whereby longer targets were named correctly significantly more often than shorter targets, even though the targets were matched for frequency and imageability. Best proposed that this reverse length effect (also illustrated by Kohn (1998) in repetition) may be due to the fact that short words tend to be similar to a greater number of other words, making them more susceptible to substitution. This hypothesis will be discussed further in the next chapter, in reference to neighbourhood effects.

A target word's level and pattern of stress may also exert an influence on its susceptibility to error. Errors tend to occur on stressed words more often than on unstressed words (Boomer & Laver, 1968; Nooteboom, 1973; Shattuck-Hufnagel, 1992) but, as mentioned earlier, this effect is confounded with word class and frequency, and may also be due to the fact that stressed words are more salient, and therefore their errors may be more easily detected (Cutler, 1981). What appears to be a more important effect of stress is the impact it has on error/target interaction. In the words of Boomer and Laver (1968), "The origin syllable and the target syllable of a slip are metrically similar, in that both are salient (stressed) or both are weak (unstressed), with salient-salient pairings predominating" (p. 7). Fromkin (1971) cautions, however, that the domain of the error is important here; Boomer and Laver's assertion is true for between-word errors, but not for within-word errors. Furthermore, in exchange errors, word-level stress patterns move with the word, while phrase-level stress patterns remain in place in the phrase. For example, in the error *nerve of a vergeous breakdown* (Fromkin, 1971), the word *nerve* takes on the secondary stress which should have been assigned to *verge*. (Note: This constraint does not discount pure stress placement errors, which do occasionally occur.)

As well as retaining the target's level of stress, numerous studies have found that errors tend to retain the overall stress pattern of the target (which also implies the preservation of target length) in both normal subjects (Bierwisch, 1981; Fay & Cutler, 1977; Shattuck-Hufnagel & Klatt, 1979) and aphasic subjects (Best, 1996; Blanken, 1990; Ellis et al., 1983; Gagnon et al., 1997; Kohn et al., 1998; Martin et al., 1994; Valdois et al., 1989). In a landmark study of on-line malapropisms produced by unimpaired speakers, Fay and Cutler (1977), compared the structural relatedness of spontaneously produced malapropisms and semantic substitutions. Whereas 75% of the semantic errors had the same number of syllables as their targets, 87% of malapropisms preserved the target's length; 82% of semantic errors shared stress patterns with their targets, relative to 98% of malapropisms. These results were interpreted as evidence that malapropisms and semantic errors, while both lexical types of error, originate at different points in the process of speech production.

Preservation of length and stress has also been noted in aphasic speech-error patterns. For example, Gagnon and colleagues (1997) found that naming responses by fluent aphasic subjects preserved the target's word shape in 74% of formal paraphasias and 70% of neologisms. It has been found that recurrent utterances in jargon aphasia, although not comparable to an identifiable target, tend to preserve normal-sounding stress patterns (Kertesz & Benson, 1970). Similarly, Buckingham et al. (1978) noted that perseverated syllabic segments retained their stress patterns. Furthermore, Blanken et al. (1988) found that responses by global aphasics, although made up entirely of neologistic speech automatisms, nevertheless showed stress patterns appropriate to the type of question being asked (wh-questions, yes/no questions, and narrative requests).

Lexical Output Biases

Two additional lexical factors warrant attention here—the effects of lexical bias and phonological facilitation. 'Lexical bias' refers to the observation that phonological errors result in real words more often than chance would predict. 'Phonological facilitation', also called the 'mixed error effect', refers to the observation that semantic errors are also phonologically related more often than chance would predict. As their labels suggest, these factors operate to bias, rather than constrain, the output of speech production. Both effects have been interpreted as evidence of the interactive nature of speech production.

Lexical Bias

Although non-word errors do occur, phonological errors have been found to result in real words more often than would be expected by chance in spontaneous speech corpora (e.g. Dell & Reich, 1981; Stemberger, 1985; but see Garrett, 1976) and in experimentally elicited errors (Baars et al., 1975). In studies of aphasic speech errors, results are more ambiguous: some researchers have found a lexical bias among aphasics' formal paraphasias (e.g. Best, 1995; Best, 1996; Blanken, 1990; Gagnon et al., 1997; Kohn & Smith, 1994a); others have not (e.g. Kohn et al., 1998; Martin et al., 1994; Nickels & Howard, 1995). In order to conclude that a lexical bias exists, it is necessary to compare the obtained rate of phonologically related real words to the chance rate. Chance is usually calculated from a pseudo-corpus of errors created by any of a variety of methods, such as randomly reassigning the error phonemes into the error slots, while respecting phonotactic constraints (Dell & Reich, 1981; Miller & Ellis, 1987; Nickels & Howard, 1995), then calculating the rate of real words produced, or randomly reassigning the word errors to targets and calculating the rate of phonological relatedness (Martin et al., 1994). Conflicting results may depend on the method of chance estimation (Stemberger, 1985), the criteria used to define phonological relatedness, or characteristics of the case studies from whom the data were collected (Nickels & Howard, 1995). It is evident, as well, that any natural corpus of formal paraphasias will inevitably contain errors of *both* phonological and lexical origin, and that the relative proportions of these may give rise to such discrepant results across studies.

Investigators have also pointed to other lexical influences on formal paraphasias to shore up findings of a lexical bias, such as a frequency effect or the preservation of grammatical class (Blanken, 1990; Gagnon et al., 1997; Martin et al., 1994). Another factor that has been called upon to identify a lexical origin for formal paraphasias is the degree of phonological relatedness. The observation that formal paraphasias (i.e. real-word phonemic paraphasias) tend to have less phonological overlap with their targets than target-related neologisms (i.e. non-word phonemic paraphasias) is interpreted as evidence that formals are errors of lexical selection, whereas target-related neologisms are errors of phonological encoding (Best, 1996; Gagnon et al., 1997; Kohn & Smith, 1994a; Martin et al., 1994).

On the other hand, Nickels and Howard (1995) argue that lexical bias effects reflect a chance outcome related to the statistical probabilities of the vocabulary. They hypothesized that, because longer words have fewer phonologically similar 'neighbours', non-word errors should be produced more frequently in response to long target words, and real-word errors produced more often in response to short target words. As predicted, the proportion of non-word phonological errors made by aphasic subjects in a naming task was found to be positively correlated with the length of the target, whereas the proportion of real-word phonological errors was negatively correlated with target length. This result alone is insufficient to discount the lexical bias effect, as acknowledged by the authors. It illustrates simply that the probabilities afforded by the lexicon contribute significantly to the likelihood of a real-word error being produced, and that estimations of chance occurrence must therefore take target length into account. When the errors of two of the subjects were compared to a length-controlled pseudo-corpus of errors, no lexical bias was shown; however, the sample of errors was quite small (n=51).

Thus, lexical status appears to be preferentially preserved in errors, at least for normal speakers. It is a truism to state that this is the case for semantic errors, but it is somewhat counter-intuitive for phonological errors; why would errors created through phonemic changes retain their lexical status? The answer depends first of all on the relatively uncontroversial postulation of a phonological lexicon where structural factors can exert an effect on lexical selection errors. But it remains to be explained why such errors occur more often than not, a subject of considerable disagreement. Baars et al. (1975) proposed the operation of an 'output editor' which preferentially allows real-word errors to slip through, a concept which has persevered in theories of speech production (Baars, 1980; Buckingham, 1980; Butterworth, 1981; Garnsey & Dell, 1984; Hofmann, 1980; Levelt, 1983; Levelt, Roelofs, & Meyer, 1999; Levelt et al., 1991a; Schlenk, Huber, & Willmes, 1987). Others, however, have proposed that the lexical bias can be accounted for more parsimoniously in an interactive spreading activation model of speech production (Dell, 1985; 1986; 1988; Dell & O'Seaghdha, 1991; Dell & Reich, 1981; Dell et al., 1997b; Harley, 1984). In such models, feedback from the phoneme to the lexical level reinforces the activation of real words, whereas non-words are not represented in the lexicon, and so cannot receive such reinforcement. The existence of a lexical bias in aphasic error studies remains unresolved, and the explanation of lexical bias effects in normal error studies remains controversial.

Phonological Facilitation

Similar accounts are called upon to explain facilitative effects of phonological relatedness, which have been shown quite consistently for normal subjects in spontaneous speech (Dell & Reich, 1981; Fay & Cutler, 1977; Fromkin, 1971; Harley, 1984; 1990; Laubstein, 1999; but see Garrett, 1980) and in experimental studies (Martin et al., 1996; Martin et al., 1989; Motley & Baars, 1976a). Results for aphasic subjects, mostly from picture naming studies, are also strong (Blanken, 1990; Dell et al., 1997b; Goodglass et al., 1997; Martin et al., 1996; but see Best, 1996), but the effect is not shown by all subjects (Dell et al., 1997b).

As mentioned earlier, the phonological facilitation effect refers to the finding that phonological relatedness between errors and targets occurs more often for semantic errors than would be expected by chance. The effect does not seem to be due to the distributional properties of semantically related words, because no such relatedness effect was found for a set of synonyms used as a control corpus (Dell & Reich, 1981), nor for a set of semantic category members (Martin et al., 1996). Martin and colleagues (1996) addressed the possibility that phonological facilitation effects in semantic errors made by aphasic subjects during a naming task might be due to the perseveration of items within the set of stimuli. Comparison of perseverated to non-perseverated semantic errors showed that this was not the case. Thus, the phonological facilitation effect appears to be a true effect influencing both normal and aphasic errors.

Like the lexical bias effect, phonological facilitation is interpreted as evidence in support of interactive activation accounts of speech production. Using a paradigm in which an array of coloured objects was described (after Levelt, 1983), Martin and colleagues (1989) manipulated the set of stimuli to create opportunities for semantically related errors (S), phonologically related errors (P) and mixed errors (S+P). Relative to their respective opportunities for occurrence, S+P errors were more likely to be produced than S and P errors combined, supporting an interactive rather than additive influence of semantic and phonological relatedness (Martin et al., 1989). Harley (1990) also referred to an interactive paradigm to explain findings of phonological facilitation among naturally occurring contextual intrusion errors. He described the effect as a result of activation 'resonating' (Stemberger, 1985) between phonological and semantic lexicons, and thus mutually reinforcing items which are connected at both levels (see also Dell, 1985; Dell & Reich, 1981; Dell et al., 1997b; Harley, 1993b). As with the lexical bias, however, a pre-articulatory editor is also able to explain the results by proposing that more closely related errors are more likely to slip through the editor's filtering function (Butterworth, 1981; Garnsey & Dell, 1984; Levelt, 1983; Motley & Baars, 1976a).

Summary

The susceptibility of words to error production is affected by a number of characteristics of the way in which lexical items are stored. Words which occur more frequently in the language appear to be more resistant to error, at least for non-brain-damaged speakers. The inconsistency of frequency effects in aphasic speech-error studies illustrates that even common words are vulnerable to error, at least in some aphasic patients, a finding which accords with clinical observations. Importantly, the frequency effects observed in error studies reflect frequencies of phonological form rather than meaning. Also affecting susceptibility are the targets' grammatical class, length and stress pattern, although it should be kept in mind that these factors are all confounded with frequency to some extent.

The outcomes of errors illustrate that the grammatical class, stress pattern, and length of the target all tend to be preserved, in both normal and aphasic errors, although these findings represent statistical probabilities rather than absolute constraints. It is also statistically more likely than chance that errors will be real words, and will be both semantically and phonologically related, suggesting that the stages of speech production proceed interactively. The statistical likelihood that errors will be more frequent than their targets has not been found consistently, perhaps in part because of the multiple constraints limiting the error's outcome. Aphasic errors do show a greater tendency than normal errors to be higher in frequency than their targets, but this observation may be partially related to the methods of analysis used. Frequency effects on outcome are shown when the average frequency of occurrence of a speech sample is compared to the normal distribution of the lexicon (e.g. Code, 1982; Howes, 1964), but this may reflect the use of empty phrases and circumlocutions, rather than specific error/target differences. Thus, different findings for normal and aphasic errors may be due in part to a lack of comparability in the methodologies used.

Syllabic Factors

In addition to preserving structural characteristics at the lexical level, error production is also influenced by structural characteristics at the syllabic level. In combination with evidence from linguistic theory, speech error studies have contributed to the establishment of the psychological reality of syllabic and sub-syllabic units. Levelt and Wheeldon (1994) found an effect of syllable frequency, independent of lexeme frequency, on picture-naming latencies, and proposed that many over-learned syllables are retrieved directly, as gestural scores from a mental syllabary (see also Sussman, 1984). The importance of the syllable as a structural framework for phonological encoding was also illustrated by Dell et al. (1993) in a parallel distributed processing simulation model, which produced a strong negative correlation between the probability of error production and the frequencies of the syllable-types encoded into the model.

In addition, studying the way in which words are broken up in the formation of syllabic intrusions (i.e. blends, MacKay, 1972), and the interaction of segments in contextual errors (e.g. Laubstein, 1987) has provided support for the existence of syllables as representational units, which in turn are composed of a binary division into onsets and rhymes (which dominate peaks and codas), or of a ternary division into onsets, peaks, and codas. Observed error constraints substantiate the notion that syllable structures exist as abstract schemas, or frames, which are filled by appropriate segments (Bierwisch, 1981; Levelt & Wheeldon, 1994; Sevald et al., 1995; Stemberger, 1990; Sussman, 1984; but see Dell et al., 1993). Strong evidence comes from the consistent finding that consonants (which fulfill onset and coda functions) interact only with other consonants, while vowels (which form the peak of the syllable) interact only with vowels (Fromkin, 1971; MacKay, 1970a; Shattuck-Hufnagel & Klatt, 1979). Fay and Cutler (1977) did report a significant number of consonant-vowel interactions in their corpus of malapropisms, but this finding has no bearing on the C/V category constraint if one accepts that malapropisms are lexical selection errors. Evidence also illustrates that the peak and coda are more likely to participate together in an error than the onset and peak, providing support for the psychological reality of the rhyme as a unit (MacKay, 1970a; 1972; Nooteboom, 1973; Shattuck-Hufnagel, 1983; but see Laubstein, 1987).

Syllable Position Constraints

Not only do segmental errors respect their syllabic category, they also respect their syllabic position (obviously these are inter-related factors). One of the most consistent and informative constraints observed in normal speech errors is the preservation of syllable position in contextual errors (Bierwisch, 1981; Boomer & Laver, 1968; Fromkin, 1971; Laubstein, 1987; MacKay, 1970a; Nooteboom, 1973; Stemberger, 1982b). (Exceptions do occur, however, in examples of within-word metatheses: *whipser, aks* (Fromkin, 1971); *fish* > *shiff*, *puck* > *cup* (Laubstein, 1987).) Confounded with this effect is the observation that interacting segments also tend have the same level of syllabic stress (e.g. Shattuck-Hufnagel, 1992). The syllable-position/syllable-stress effect has also been found in aphasic error studies (e.g. Kohn & Smith, 1990). However, contextual phonological errors are much less common in aphasic speech than in normal speech, relative to non-contextual errors (Stemberger, 1982b; Talo, 1980), providing fewer opportunities to observe these effects. Nevertheless, syllabic position and stress have been noted to constrain the production of alliterative and assonantal stretches of perseverated neologistic jargon (Buckingham & Kertesz, 1974; Buckingham et al., 1978).

Another way in which syllable position exerts an effect on error production is in the differential susceptibility to error of segments within a syllable, particularly syllable onsets. In analyses of normal speech errors, the most significant proportion of phonological errors disrupt consonants in word-initial position (Bierwisch, 1981; Dell & Reich, 1981; Garnham et al., 1981; MacKay, 1970a; Shattuck-Hufnagel, 1987; Shattuck-Hufnagel, 1992; Shattuck-Hufnagel & Klatt, 1979). MacKay (1970a) found that both within-word and between-word reversals collected from the spontaneous speech corpus of Meringer and Mayer (1895, cited in MacKay, 1970a) occurred on syllable-initial consonants at greater than chance levels, but that word-initial reversals were more common than syllable-initial reversals. Shattuck-Hufnagel (1992) later confirmed this finding in a series of tongue-twister experiments, showing that, when stress level was controlled, word-onset consonants were twice as likely to be involved in errors as syllable-onset consonants which were not word-initial. Thus, it is not only the syllable-position which is important, but also the word-position of the segment. Shattuck-Hufnagel (1987) proposed that word-onsets must be afforded a special status in models of phonological encoding. Alternatively, Dell and colleagues (1993) suggested that the vulnerability of word onsets may be related to their relative lack of predictability.

Unlike normal subjects, aphasic subjects have been observed to make *fewer* errors on onsets than on segments in other positions (Gagnon & Schwartz, 1997; Gagnon et al., 1997; Kohn, 1989; Kohn & Smith, 1990; Martin et al., 1994; Martin et al., 1996; Romani & Calabrese, 1998; but see Blanken, 1990). However, this apparent difference between normal and aphasic speakers probably has more to do with the types of errors studied in each case. As Meyer (1992) pointed out (referring to normal speech errors), "what makes the word-onset effect particularly intriguing is that in sound errors word onsets are particularly vulnerable, whereas in malapropisms and TOT [tip-of-the-tongue] states they are more likely to be correct" (p. 188). This was true for the malapropisms in Fay and Cutler's (1977) study, and for the majority of formal paraphasias produced by aphasics (Gagnon & Schwartz, 1997; Gagnon et al., 1997; Martin et al., 1994; but again, not in Blanken, 1990), and was also found to be true of semantic errors produced by both normal and aphasic subjects (Martin et al., 1996). Furthermore, perceptual studies have shown that word onsets are more salient, so word-onset errors are probably more detectable (Meyer, 1992). Although this might explain the apparent word-onset susceptibility in normal speech errors, it would be difficult to reconcile with findings of relatively preserved onsets in aphasic speech production. A final source of discrepancy among studies may relate to the language-dependency of onset structures.

Thus, there is still an unresolved discrepancy between the sound errors of normal subjects, which tend to disrupt onsets, and the sound errors of aphasic subjects, which tend to preserve onsets (Kohn, 1989; Kohn & Smith, 1990; Martin et al., 1994; Romani & Calabrese, 1998). Again, the critical difference may reside in the distributions of contextual and non-contextual errors in the two populations. In addition, the heterogeneity of aphasic deficits in case studies clearly contributes to the differential findings. Kohn and Smith (1995) found that onsets were preferentially preserved in only three of their six fluent aphasic subjects, who also produced many fragment errors. Because fragments contain only word-initial segments, their abundance may inflate the rate of onset preservation. (Also note that there is some overlap of subjects across Kohn's studies; subject CM appeared in the studies of Kohn (1989), and Kohn and Smith (1990; 1995).) The authors concluded from these results that, because these three subjects showed a deterioration in performance throughout the word (along with other phonological evidence), they had deficits in phonological planning, whereas their other subjects had deficits in phonological activation.

Similar differences have been shown among clinical sub-types of aphasia. Onsets have been found to be relatively more difficult for Broca's aphasics (Trost & Canter, 1974), but relatively less difficult for Wernicke's and conduction aphasics (Burns & Canter, 1977). This group difference, which was confirmed in a reanalysis of the data by Canter, Trost and Burns (1985), was at least partly attributed to an apraxic component in the Broca's aphasics, making it more difficult for them to initiate articulation accurately, and thus disrupting onsets more frequently (Canter et al., 1985). However, these findings are somewhat contradictory to Kohn's finding that onsets were *more* difficult in some fluent aphasics. It may come down, again, to a lack of comparability in the types of errors across studies; it may be, for example, that the onset preservation effect in the fluent group of Burns and Canter is due to an over-representation of semantic errors by these subjects. What is clear from these studies is that no definitive conclusions can be drawn about onset constraints in aphasia without carefully controlling the types of errors being compared (in particular, whether they take place at the lexical or phonological level) and whether or not they are contextually determined. Furthermore, it is important to take into account the characteristics of the aphasic subjects being tested and their levels of speech production deficit.

Syllable Markedness

According to Nespoulous and colleagues, various definitions of 'markedness' from different domains, such as historical linguistics, physiology and perception, and language development, have given the concept of markedness "a somewhat heterogeneous flavor, with frequent overlaps" across domains (1984, p. 204). Syllable markedness refers to a combination of factors—frequency within a language, universality across languages, length and complexity—which together create a continuum of syllable types (Nespoulous et al., 1984). For example, Favreau and colleagues (1990) defined a hierarchy of markedness in their bi-syllabic stimuli, in which CV-CV stimuli were the least marked, and CV-CCVC were among the most marked. Because studies of normal speech errors have focused primarily on the preservation of syllabic structures in contextual phonological errors, there is little speech-error data that speaks to markedness; if syllable structure is preserved, there is no change in markedness. Aphasic speech production, however, has revealed an influence of markedness in the creation of errors.

In several investigations, aphasic speech errors have been shown to reduce the markedness of the syllable structure. Favreau et al. (1990) manipulated the syllable markedness of word and non-word stimuli and compared the numbers of errors made by aphasic subjects in a repetition task. Initial results indicated that unmarked syllables resulted in fewer errors, and that the majority of errors reduced syllable markedness, but further analysis revealed that this effect was related to the length of the stimuli. In other studies, similar effects have been found. Consonant omission, especially in cluster reduction, has been noted to be the most common phonological process in a number of aphasic error corpora (e.g. Béland, Paradis, & Bois, 1993; Parsons, Lambier, & Miller, 1988, reanalyzing errors from several previous studies), and consonant clusters have been found to be more error-prone than singletons (Blumstein, 1973a; Stemberger & Treiman, 1986; Trost & Canter, 1974). In recurrent utterances, CV syllables are most common, and clusters are rare (Code, 1982). Thus, length is clearly a factor in determining syllable markedness.

Other research, however, illustrates that markedness extends beyond the number of segments. Many aphasic errors increase syllabic complexity (i.e. number of segments), while decreasing syllabic markedness (Béland et al., 1990; Béland et al., 1993; Kohn & Smith, 1994a). Consonants may be added to create onsets, resulting in the least-marked CV syllable, as in *elephant* > /vɛləfənt/ (Kohn & Smith, 1994a); consonants may be added between vowels, as in *poème* > /polɛm/ (Béland et al., 1990); and vowels may be added within consonant clusters, as in *strie* > /sœtri/ (Béland et al., 1990), and *pumpkin* > /pʌpəkɪn/ (Kohn & Smith, 1994a). Martin et al.

(1975) found that additions were more likely in the repetition of CV stimuli, while omissions were more likely in CCVCC stimuli, with the result that aphasic subjects tended to produce the canonical CVC word form.

One of the most important factors in determining markedness is the concept of sonority sequencing, which describes the optimal order of sounds in a syllable in terms of perceptual salience and articulatory openness (Romani & Calabrese, 1998). Preferred syllables are those with a maximal sonority differential between onset and peak (e.g. stop + vowel), and a minimal sonority differential between peak and coda (e.g. vowel + nasal). Sonority sequencing also dictates which sequences of phonemes are phonotactically impermissible. Syllables in aphasic non-word errors have been noted to adhere to the principles of sonority sequencing, both for target-related neologisms (Christman, 1994) and recurrent, non-target-related utterances (Code & Ball, 1994). Where targets were identifiable, errors were observed to maintain the sonority profile of the target syllables most of the time; when changes in sonority did occur, they increased the sonority profile (i.e. reduced the complexity) of the target (Christman, 1994; Kohn et al., 1998).

The general tendency to decrease markedness may not be consistent across all types of aphasia, or all types of errors. It has been observed, for example, that subjects with Broca's aphasia are more likely to reduce syllable markedness through cluster reduction than are conduction aphasics (Nespoulous et al., 1984; see also Bastiaanse et al., 1994; Burns & Canter, 1977; Trost & Canter, 1974). Similarly, Kohn and Smith (1994a) found a reduction in markedness in the phonological errors of one of their two subjects, contributing to the diagnosis of a deficit in lexical-phonological activation as opposed to phonological planning. Gagnon et al. (1996) noted that remote neologisms tended to show a reduction in markedness, while target-related neologisms were equally likely to create more marked and less marked syllable structures. Non-contextual errors also show a greater tendency to reduce markedness than do contextual errors (Christman, 1994). Thus, the creation of less marked syllable structures in paraphasias may be related to factors other than syllabic structure preferences, such as contextual influences or, as noted by Kohn (1984), the use of a high-frequency syllable such as *ing*, which also has salience by virtue of its morphological status.

These results seem to suggest that constraints such as sonority sequencing and markedness reduction are revealed in the absence of other overwhelming influences. Béland (Béland et al., 1990; Béland et al., 1993) interprets her results as indicative of 'repair strategies' which are informed by an implicit knowledge of phonology, and operate in aphasia to circumvent deficits in phonological production processes. That is, syllable markedness is reduced when aphasic speakers mistakenly perceive constraint violations in complex syllables and attempt to repair them (Béland et al., 1993). Another sort of compensatory 'strategy' has been proposed in the form of a random generator (e.g. Buckingham, 1981, 1990b; Butterworth, 1979); both tend to create relatively unmarked structures, except that Béland's relies on phonological rules rather than 'random volleys' to produce neologisms. As Christman (1994) speculates, "it is logical that a damaged system (in the interest of self-preservation) might revert to production of its least challenging product when stressed" (p. 114).

Summary

Error studies have provided evidence for the psychological reality of syllabic and sub-syllabic units. Syllabic frequency effects suggest that there may even be a separate store for syllables, or syllabary; this is most likely for highly frequent syllables (Levelt & Wheeldon, 1994). At the least, syllabic units are represented at separate levels of the lexicon, as shown by their differential susceptibility to error. In particular, syllable onsets are more likely to be disrupted than syllable rhymes in normal speech errors, but more likely to be preserved in aphasic speech errors. Although this difference may reflect a true deficit in producing syllabic rhymes in some types of aphasia (Kohn & Smith, 1995), it may also be due to differences in the types of errors represented in normal and aphasic corpora.

Syllabic constituents also play a role in the outcome of errors. In the majority of contextual speech errors, interacting elements share syllabic position and syllabic stress level. Because contextual errors are more common in normal than in aphasic speech, this effect is observed largely in normal speech-error studies. On the other hand, effects of syllable markedness on error outcome are observed most frequently in aphasic errors. When syllabic structure is not preserved, aphasic errors show a tendency to simplify syllables by creating less marked syllables, such as syllables with fewer phonemes, or syllables with a preferred sonority profile.

Phonological Factors

As for syllabic representations, speech error data have provided compelling evidence for the psychological reality of phonemes as primary units of speech production. Among the phonological errors collected in spontaneous speech corpora of both normal and aphasic speakers, single phoneme errors constitute the most frequent type of error (Béland et al., 1990; Blumstein, 1973a; Boomer & Laver, 1968; Dell & Reich, 1981; Fromkin, 1971; Garnham et al., 1981; Nooteboom, 1973; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1985). However, there are some conflicting findings concerning exactly what constitutes a phoneme. For example, Fromkin (1971) concluded from her observations that clusters are made up of discrete phones which may act independently, and that /ŋ/ also is divisible into two component phonemes /n/ and /g/, but that affricates are indissoluble. Others, however, have claimed that affricates may be broken up (e.g. Shattuck-Hufnagel & Klatt, 1979), but that consonant clusters rarely are (at least in exchange errors, MacKay, 1970a).

Speech errors have also provided somewhat contradictory evidence about the psychological reality of distinctive features. On one hand, substitutions are the most common type of single-phoneme error and, in several studies, the majority of substitutions involve a change of only one feature (Blumstein, 1973a; Romani & Calabrese, 1998; Stemberger, 1982b; Trost & Canter, 1974; but see Burns & Canter, 1977). In other studies, the phonetic overlap between errors and targets is at least greater than chance (Burns & Canter, 1977; Fay & Cutler, 1977; Fromkin, 1971; Green, 1969; Lecours & Lhermitte, 1969; MacKay, 1970a; Nooteboom, 1973; but see Levitt & Healy, 1985; Boomer & Laver, 1968), suggesting that it is features which are being substituted rather than whole phonemes. On the other hand, only a few examples can be confidently attributed to the feature level. As discussed in the last chapter, errors such as *glear plue sky* (Fromkin, 1971) are more parsimoniously explained as feature exchanges than as independent phoneme exchanges. Such examples, however, occur rarely—by one estimate, about fifty times less often than segmental exchanges (Shattuck-Hufnagel, 1983). It has been argued that, if features exist as truly independent units, the rate of single-feature errors would be much higher (Shattuck-Hufnagel, 1983; Shattuck-Hufnagel & Klatt, 1979). Alternatively, the rarity of independent feature involvement in errors has been accounted for by their closed-class status (Dell, 1990). Features cannot recombine to form new phonemes the way phonemes can to form new words; they "cannot exist except as properties of larger segments" (Fromkin, 1971, p. 37). Moreover, the relationship between the incidence of errors and phonetic similarity depends on the feature system used (Buckingham, 1987; Lecours & Caplan, 1975; Levitt & Healy, 1985). The status of distinctive features as independent units remains unresolved.

Phoneme Frequency

Like lexical frequency and syllable frequency, the frequency with which a phoneme occurs in the language may influence its involvement in error productions. A few studies have shown a negative correlation between phoneme frequency and error incidence (e.g. Blumstein, 1973b; Levitt & Healy, 1985; Shattuck-Hufnagel & Klatt, 1979; Trost & Canter, 1974), but the effect of frequency on error outcome is unclear. While more common phonemes tend to be produced more accurately, they also tend to occur more often in errors (Shattuck-Hufnagel & Klatt, 1979; but see Levitt & Healy, 1985). Levitt and Healy (1985) found that the incidence of phonemes occurring in intrusions (i.e. errors) was not related to phoneme frequency counts, but that intrusion phonemes were more often higher in frequency than their targets. Shattuck-Hufnagel (1979) cited a significant positive correlation between error involvement in targets and intrusions as evidence that frequency does not influence error outcome, but suggested that frequency counts using word-onsets only might be more appropriate, since the onsets were most frequently involved in errors. Furthermore, she found some anomalous tendencies, in which the alveolars /s/ and /t/ were replaced by the less frequent palatals /ʃ/ and /tʃ/. Stemberger (1991) extended this result to experimentally elicited contextual errors, demonstrating such an 'anti-frequency' effect in other substitutions. He hypothesized that this bias was due to the underspecification of the phonemes relative to their intrusions; when these segments compete, the more fully specified segment wins out.

Aphasic errors tend to retain the phoneme frequency distribution of the language (Blumstein, 1973a; Green, 1969), although in severe cases, the distribution of phonemes may be restricted. For example, Code (1982) distinguished real-word from non-word recurrent utterances; the real words reflected a normal phoneme distribution, making use of 40 out of a possible 44 phonemes, whereas the non-words were made up of only 21 different phonemes. Butterworth (1979) found that the distribution of initial phonemes from non-target-related neologisms differed significantly from the distribution of initial phonemes from content words, verbal paraphasias, and target-related neologisms. In fact, they were generally *lower* in frequency, which the author interpreted as evidence of the random selection of phonemes during production of these 'device-generated' neologisms.

Phoneme Markedness

Like syllables, phonemes also differ in their degree of markedness. Whereas syllable markedness is related to the way in which segments are combined, phoneme markedness is related to their feature composition. There is a hierarchy of feature specification which places sonorance at the top, followed by manner of articulation and voicing, with place of articulation at the bottom (see Béland, 1998). Segments are either marked or unmarked for each feature. At the top of the hierarchy, sonority plays a large role in determining markedness but, unlike the gradient of sonority in syllable structure, segmental sonority is a binary feature, like all the other features, separating obstruents from sonorants (but see Bastiaanse et al., 1994).

This hierarchy is reflected in the incidence of different types of segmental substitutions. Findings that vowel errors are far more rare than consonant errors (e.g.

Baum & Slatkovsky, 1993; Béland, 1998; Blumstein, 1973a; Burns & Canter, 1977; Garnham et al., 1981; Green, 1969; Kohn et al., 1998; Monoi et al., 1983; Shattuck-Hufnagel & Klatt, 1979; Sussman, 1984; Trost & Canter, 1974) reflect the primacy of the sonorant feature. (But see Bastiaanse et al., 1994; Christman, 1994; Kohn & Smith, 1990; and Monoi et al., 1983 for an interesting lack of sonority preservation in conduction aphasia.) Among consonant errors, place of articulation is the most frequently disrupted feature, followed by manner, nasality and voicing, in various orders (Burns & Canter, 1977; Green, 1969; Kohn et al., 1998; Kohn et al., 1995; Trost & Canter, 1974; see also Buckingham, 1987 for a review). Thus, markedness appears to have a strong effect on the relative vulnerability of particular segments to error production.

As in analyses of syllable structure in errors, the influence of phoneme markedness on error outcome is less clear-cut. Motley and Baars (1975) found no markedness effects in elicited spoonerisms, but suspected that the range of markedness contrasts in their stimuli was too narrow to show an effect. Shattuck-Hufnagel and Klatt (1980) compared three models of phoneme substitution using a confusion matrix of target and intrusion segments. In the first model, segments are substituted at random (supported by data from Boomer & Laver, 1968); in the second model, the markedness model, 'stronger' segments replace 'weaker' segments; in the third model, segment intrusions are conditioned by their availability within the planning frame, and their phonetic similarity to the target. As in a previous study (Shattuck-Hufnagel & Klatt, 1979), the confusion matrix was largely symmetrical, indicating that segments were equally likely to appear as intrusions and targets, but the segment substitutions showed contextual influences and target-error similarity, supporting the third model. In a similar analysis, Stemberger (1991) found that less marked items are actually more prone to error, and hypothesized that where null specification competes with specification, the specified segment receives a greater amount of activation and is, thus, more likely to be selected. (These results parallel the results reported earlier regarding phoneme frequency effects in this study, no doubt because of the close relationship between frequency and markedness.)

Although there appears to be no effect of markedness, or possibly even an 'anti-markedness' effect on the outcome of normal errors, Blumstein (1973a) found that substitution errors tended to replace marked with unmarked segments in Broca's, Wernicke's and conduction aphasics (though the difference was not significantly different from chance for conduction aphasics). Similarly, a case study by Romani and Calabrese (1998) showed that 55% of single-feature consonant substitutions resulted in a less marked segment, and 45% in a more marked segment. However, given that this difference is quite small, they concluded that syllabic complexity is a more important factor in determining the error outcome than segmental markedness (see also Buckingham, 1980; Kohn et al., 1998).

Phonotactic Constraints

Even in such apparently randomly created errors as abstruse neologisms, the outcome of the error is almost always a permissible string of phonemes, according to universal and language-specific phoneme sequencing rules (Boomer & Laver, 1968; Buckingham & Kertesz, 1974; Buckingham, 1980; Buckingham, 1987; Buckingham, 1990; Butterworth, 1979; Christman, 1994; Code, 1982; Fromkin, 1971; Green, 1969; Lecours & Rouillon, 1976; Lecours & Lhermitte, 1969; Sussman, 1984). In other words, phoneme sequences that do not occur in real words in a given language will also not occur in errors in that language. As stated by Wells in 1951, this is the 'first law of tongue slips': "a slip of the tongue is practically always a phonetically possible noise" (cited in Boomer & Laver, 1968, p. 7).

Some researchers have stated the phonotactic constraint very strongly (e.g. Fromkin, 1971; Sussman, 1984), although there is now evidence that it may not be as inviolable as our perceptions would lead us to think. As discussed in the last chapter, the greatest impediment to an accurate assessment of phonotactics in normal speech, especially from samples of spontaneous speech, is our perceptual bias to filter out any violations (Cutler, 1981; Meyer, 1992; Mowrey & MacKay, 1990). In a computer simulation of error production, Dell et al. (1993) found that the model produced word-final syllables which, while phonotactically legal, do not occur in English (e.g. /fɪ/, /mɛ/). While acknowledging the possibly that these errors come from a 'bug' in the model, the authors suggested that such errors in normal speech "may indeed occur and be incorrectly coded as cutoff words, rather than phonotactic violations" (p. 176). According to Buckingham (1980), "under conditions of rapid speech and with closer analysis some of these constraints will be broken on the part of the speaker", but "in many instances the hearer unconsciously reanalyzes the form according to the constraints" (p. 209). Even more problematic is making the distinction between errors of phonological selection and errors of articulation (Buckingham & Yule, 1987), both of which can occur in fluent and non-fluent aphasics (Blumstein, 1973a; Blumstein et al., 1980; Buckingham & Yule, 1987). Because articulatory errors occur following the stages of phonemic selection and sequencing, they may display phonotactic violations.

Aside from errors attributed to motor speech deficits, assuming they can be distinguished from phoneme selection errors, it is generally agreed that phonotactic constraints are respected in aphasic speech errors, even in abstruse neologisms

(Butterworth, 1979), recurrent automatisms (Code, 1982) and glossolalic output (Lecours, 1982). In fact, Béland and colleagues claim that aphasic speech is even more tightly constrained than normal speech, causing them to 'repair' perceived violations by replacing them with less marked structures. Buckingham (1987) advanced a similar hypothesis to account for errors promoted by repeated phonemes in the stimulus—a 'hyper-sensitive' error monitor will tend to 'check off' repeated phonemes, even when they are required, producing errors of simplification.

The phonotactic constraint is not simply a conclusion drawn from null findings; it is an active process similar to the repair strategies described by Béland (see above). This is illustrated in example 5d in Table 1, in which *play the victor* is produced as *flay the pictor** (Fromkin, 1971). The /v/ transposed from *victor* changes to /f/ in combination with the /l/ from *play* to prevent the formation *vlay**, a sequence which, while respecting sonority profiles, is nevertheless phonotactically illegal in English. A similar phenomenon occurs with morphophonemic accommodation (Bierwisch, 1981; Fromkin, 1971; Garrett, 1980). In example 6c, *tab stops* > *tap stobs**; (Fromkin, 1971), the plural morpheme /s/ in *stops* changes to /z/ in the error *stobs**; in example 5e, *an eating marathon* > *a meeting arathon**; (Fromkin, 1971), the indefinite article *an* becomes *a* to accommodate to its new environment in front of *meeting*. Thus, as Fromkin explained, "phonological constraints, when learned, become behavioral constraints which occur AFTER the segmental transpositions occur" (p. 41).

Summary

Both the frequency and the markedness value of phonemes appear to contribute to their error susceptibility, but have a questionable role to play in determining the outcome of normal errors. Aphasic errors, however, seem to be more strongly influenced by phoneme frequency and markedness. Phonotactic constraints apply solely to error outcome. Whereas phonotactics and morphotactics determine which patterns are permissible in the language, factors such as phoneme frequency and phoneme markedness determine which patterns are 'preferred'. Thus phonotactic rules act as strong (though not inviolable) constraints on error outcome, while frequency and markedness reflect statistical tendencies in the language. As statistical probabilities, it appears that they are often over-ridden by other, possibly stronger, influences, such as lexical-level constraints, syllabic structure constraints, or the availability of segments within the context of the utterance.

Contextual Factors or 'Availability'

As is evident from the preceding sections, many of the constraints on error production exert their influence on the interaction between targeted and intruding elements, and in many cases (especially in error elicitation experiments) the intruding elements come from within the linguistic context of the utterance. To review: functional and structural similarities have been noted between interacting error elements. For word-level substitutions and exchanges, the grammatical and morphological characteristics of the target are almost always preserved—content words interact with other content words; nouns substitute for nouns, verbs for verbs, and roots for roots (Buckingham, 1980; Butterworth, 1979; Dell, 1990; Fay & Cutler, 1977; Garrett, 1980). At sub-lexical levels, similar constraints are observed—vowels substitute for vowels and consonants for consonants; stressed syllables interact with other stressed syllables; word onsets substitute for word onsets and rhymes for rhymes (Fromkin, 1971; Garrett, 1980; Shattuck-Hufnagel, 1992). Segmental errors have also been found to be phonetically similar more often than dictated by chance in both normal (Fromkin, 1971;

Shattuck-Hufnagel, 1979) and aphasic errors (Blumstein, 1973b; Burns & Canter, 1977; Trost & Canter, 1974; but see Ellis, 1985).

The availability of such similarities within the same utterance can promote their interaction and, thus, the creation of contextual errors. This is the principle exploited by tongue-twisters and many other error elicitation techniques. Context can also promote errors by the availability of *preferred* items within the utterance; an element may be replaced by one that is more frequent, less marked or less complex or, alternatively, more fully specified. In this section, a brief review is provided of research concerning the influences of linguistic context (in particular the phonemic environment) and non-linguistic context. But first, the difficulties associated with defining the contextual domain of an utterance are discussed.

Contextual Domain

Defining what constitutes the 'environment' or the 'availability' of segments is not an easy matter. The domain over which contextual errors may interact has generally been assumed to be about the size of a phrase or clause (e.g. Boomer & Laver, 1968; Buckingham, 1987; Christman, 1994; Dell, 1986; Fromkin, 1993; Garrett, 1975; MacKay, 1970a; Shattuck-Hufnagel, 1992). Because different units have planning frames of different sizes, however, the domain depends on the level of the error, and thus the task used to elicit the errors; segmental interactions tend to be restricted to the same clause, whereas words can interact across two clauses (Dell, 1986; Garrett, 1975; Pate et al., 1987; Shattuck-Hufnagel, 1992). Different domains may also apply to anticipatory and perseveratory errors. As Schwartz et al. (1994) point out:

> When contextual influences are examined, the results are influenced by how large a window one applies, how symmetrically applied it is in the forward (anticipatory) and backward (perseveratory) direction, and most important, what constraints are imposed on the source-error pairs. (p. 59)

Furthermore, Buckingham (1980) suggests that aphasic errors, particularly perseveratory errors, can span a wider domain than normal errors. As mentioned in the previous chapter, contextual influence may be more appropriately viewed as graded, rather than an all-or-none variable (Dell et al., 1993). Wheeler and Touretsky (1997) further suggest that all errors might be considered contextual, "if one takes a broad enough view of 'context'" (p. 160). That is, every level of speech production has its own 'context' defined by the availability of competing intentions (see also Baars, 1980). During sentence formulation, the utterance provides a syntagmatic context; during lexical selection, the semantic and phonological lexicons operate as paradigmatic contexts.

Phonemic Environment

Contextual influences can extend beyond the intruding segment to other segments in the environment of the target. One aspect of the phonemic environment which has been shown to influence error susceptibility is the transitional probabilities of the phonemes. As Motley and Baars (1975) noted, "spoonerisms are facilitated when one of the phonemes in a phoneme string destined for articulation enjoys a greater probability of occurrence in an earlier-than-intended context than does the phoneme originally intended for that context" (p. 360). Another environmental effect that has been observed is that errors tend to cluster together; thus, in the environment of an error, the probability of making another error is increased (Ferber, 1991). This tendency is exaggerated in some aphasic patients who experience a sort of 'noise build-up' throughout a speech task (Brookshire, 1992).

A number of studies have revealed effects of repetition, whereby the probability of error increases when phonemes are repeated in the utterance (Dell, 1984; Dell, 1988; Fromkin, 1971; Garrett, 1980; Lecours & Lhermitte, 1969; MacKay, 1970a;

Nooteboom, 1973; Shattuck-Hufnagel, 1992; Shattuck-Hufnagel & Klatt, 1979; Stemberger, 1990; Stemberger, 1991; but see Sevald & Dell, 1994 for some inhibitory effects). One type of repetition effect is exploited by tongue-twisters. To illustrate, the replacement of /k/ by /p/ in *Spanish speaping* people* (Fromkin, 1971) is facilitated by the three other occurrences of /p/ in syllable-onset position in the phrase. A different mechanism is observed in the 'repeated phoneme effect' illustrated by Dell (1984), in which the repeated phoneme is not involved in the error, but creates an environment that encourages the adjacent phonemes to slip. For example, the exchange *heft* lemisphere** is facilitated by the fact that the exchanging onsets are both followed by the same vowel. In addition to shared phonemic content, shared 'wordshape' (i.e. syllable structure and stress pattern) can also contribute to error production, as demonstrated by Stemberger (1990) in errors of cluster creation (see also Sevald et al., 1995).

Error outcomes are obviously conditioned by the nature of the intruding segment, but outcome may also be influenced by the surrounding phonemes. The adaptation of intruding phonemes to their new environments during phonotactic and morphological accommodation, described earlier in the chapter, is an illustration of this. Kohn and her colleagues (1995) revealed yet another process of phonemic adaptation of errors in aphasic speech. Consonant harmony, a process of copying features from the phonological environment (defined as the target word in this case), was found to exert a significant influence (at least for the voicing feature) in determining the outcome of phonologically related errors by fluent aphasic subjects. For example, in the error *vest* > *fes**, the initial phoneme takes on the voicing value of the /s/. The authors concluded that the process operates as a sort of compensatory strategy: "fluent aphasics may draw upon the particular phonological rule system of their language as a compensatory mechanism to reconstruct utterances based on faulty lexical-phonological information" (Kohn et al., 1995, p. 755).

Non-Linguistic Context

Certain non-linguistic aspects of the speaking situation may also influence error production. Although these will not be discussed in detail, a few factors which are particularly relevant to aphasic error studies bear mentioning. Some studies have found different distributions of errors depending on the task used and the nature of the stimuli (Barton et al., 1969; Basso, Razzano, Faglioni, & Zanobio, 1990; Williams & Canter, 1982; but see Kohn et al., 1992). For example, Williams and Canter (1982) found that Broca's aphasics performed more accurately on a confrontation naming task than on a picture description task, whereas Wernicke's aphasics showed the opposite pattern. Dell and his colleagues (1997b) have demonstrated that practice has the effect of bringing aphasic patterns of error production closer to normal patterns, at least with a restricted set of stimuli. On the other hand, Butterworth (1992) reminds us of the lack of consistency of responses across testing sessions on any particular item. Thus, testing (and possibly retesting) on a large number of stimuli is necessary before drawing conclusions about the relationship of the errors produced to any given characteristic of the stimulus items. The attentional demands of the task may also affect lexical access abilities in aphasic subjects (e.g. Martin et al., 1975; Murray, 2000). Aphasic subjects are often highly distractable, so care must be taken to control potential external sources of 'noise' which might give rise to environmental intrusions abstruse to the experimenter.

Summary

Defining what constitutes contextual influences depends on the domain of the errors under study and the particular demands of the task. In addition to lexical and sub-lexical characteristics of the targets and errors themselves, error production is also affected by characteristics of the linguistic and non-linguistic contexts in which they occur. Target items become more susceptible to error if there are confusable items—for example, items of similar structure and function, or items which are higher in frequency or less marked—in the immediate vicinity. Moreover, items embedded in similar phonological environments (as in tongue twisters) can also be more easily induced to slip. Aspects of the non-linguistic context may also promote errors further, by stressing the processes of speech production.

Investigators have noted a trade-off between the availability of confusable items and other error-promoting factors. For example, Stemberger (1982b) found, in a study of spontaneous errors collected from normal speakers, that fewer within-word than between-word sequencing errors differed from their target by only one feature. Martin et al. (1998) analyzed perseveratory responses by an aphasic subject in a naming task, and revealed that perseverations which were neither semantically nor phonologically related to the target had more recent sources than related perseverations. Finally, Levitt and Healy (1985) showed a trade-off between availability (i.e. contextual vs non-contextual) and phoneme frequency in a study of phoneme target-intrusion interactions. They reported that "segment availability becomes increasingly important as the frequency of the intruded phoneme decreases and perhaps, to a lesser extent, as the featural similarity between the intruded and target phonemes decreases" (p. 732).

Chapter Summary

Patterns of error occurrence (and non-occurrence) have provided a valuable source of information about how linguistic structures are represented and how they interact in speech production. In many respects, findings from error studies corroborate findings from linguistic theory and from other speech-production paradigms, such as tip-of-the-tongue studies (e.g. Goodglass, Kaplan, Weintraub, & Ackerman, 1976; Harley & Bown, 1998), word games (e.g. Treiman, 1983), cueing studies (e.g. Pease & Goodglass, 1978; Spencer et al., 2000), and reaction time paradigms (e.g. Levelt et al., 1991a; Schriefers et al., 1990). The psychological reality of representations at lexical, syllabic, phonemic, and feature levels gains support from speech errors. Constraints make items at various levels more or less susceptible to error, they may influence the nature of the error produced, or they may do both. It is often unclear whether constraints are acting on what is available, or what can be produced (Stemberger, 1982b). What is clear from the preceding discussion is that no direct relationship between slipability and outcome has yet been identified. For example, a low frequency of occurrence may contribute to the inaccurate production of an item, but that does not mean that the error produced will be of higher frequency. In general, error outcomes appear to be less predictable than might be expected, given the factors promoting and limiting their occurrence.

Although the previous discussion has been organized according to linguistic level, there are similarities in constraints across the levels. Frequency effects occur at lexical and phoneme levels, and possibly also at the syllable level; markedness influences exist at both syllable and feature levels. The evidence for syllabic and feature representations is somewhat paradoxical, what Dell (1986) calls the 'units problem'. That is, although these structures participate rarely as independent units in errors, they are subject to similarity constraints which support their existence. 'Slot-filler mechanisms', that is, the formulation of a planning frame into which units are subsequently inserted (described more fully in the next chapter), also appear to operate at different levels. Words are inserted into sentence frames, and phonemes are inserted into syllabic frames; the misfiring of these mechanisms gives rise to characteristic errors. The slot-filler paradigm illustrates another parallel that can be drawn across different linguistic levels—the distinction between open and closed classes. Because open class constituents (e.g. content words and phonemes) must be selected and ordered during speech production, whereas closed class constituents (e.g. function words and features) are automatically retrieved during frame construction, the two types of words show different error influences (Dell, 1990; Garrett, 1980).

In addition to these parallels, there are also many interactions and confounds among the constraints at different levels. For example, frequency and markedness are inter-related, as are syllable complexity and markedness. Grammatical class is confounded with word stress, and syllabic position constraints are confounded with syllabic stress constraints. Many of the elements observed to participate most often in errors (e.g. content words, stressed syllables, initial consonants) are those which are also most perceptually salient. Finally, constraints appear to 'trade off' in different contexts, such that a strong constraint in a given context might obscure or even over-rule the action of another operative constraint. Baars (1992a) hypothesized that the production of errors is a reflection of a three-way trade-off in the speech production system among the goals of flexibility, speed, and accuracy. Some aspects of speech production are rigid and automatized and less prone to error, whereas other aspects must retain the flexibility required in a language system. "The more choices we have," Baars notes, "the slower our response will be" (p. 20), and the more choices we have, the more errors we will make.

None of the constraints, then, are absolute. According to Martin and colleagues (1989): "Constraints on error occurrence reflect those properties of speech errors which appear to be invariant... Probabilistic influences refer to effects of linguistic variables which may not be the primary source of errors, but increase the likelihood of their occurrence" (p. 463). It is doubtful, however, that any of the properties of speech errors discussed here are truly 'invariant'. In contrast to Martin et al., Kohn and colleagues (1995) consider constraints to operate along a continuum, from 'rule-oriented' to 'random'. Some constraints are evidently stronger than others, but much of the variability in the application of a constraint can be accounted for by differences in error collection techniques, error classification systems, and artifacts of the analyses.

Although aphasic errors seem to be subject to the same constraints as normal errors, some differences have been noted in the patterns of errors produced by the two populations. In corpora of spontaneous errors, normal speakers produce a relatively greater proportion of contextual compared to non-contextual errors than do aphasic speakers (e.g. Stemberger, 1982b; Talo, 1980). For those aphasic errors which do have a contextual source, the intrusion and the target come from the same word more often than in normal errors (e.g. Pate et al., 1987; Schwartz et al., 1994). Anticipations are more common than perseverations in normal speakers, whereas the reverse is true for aphasic speakers (Dell et al., 1997b; Schwartz et al., 1994). The susceptibility of word onsets seen in normal errors (e.g. Shattuck-Hufnagel, 1987) is not reliably observed in aphasic errors (e.g. Kohn & Smith, 1990). Some aphasic subjects show a greater involvement of vowels in errors than do normal subjects (e.g. Christman, 1994; Kohn & Smith, 1990). Talo (1980), and many others, have also noted that normal speakers are usually more aware of their errors. Thus, they may be more effective at on-line monitoring and repairing errors than most aphasic speakers. Schwartz and her colleagues (1994) note that observed differences between aphasic and normal error patterns may also be due, in part, to the use of different criteria in defining error categories. In aphasic error studies, of which many are case studies (and many of these are unusual cases, e.g. Best, 1996; Blanken, 1990), a considerable amount of variability may also be attributed to the heterogeneity of the aphasic population. But the type and domain of the errors being studied appear to be more important variables than clinical sub-type of aphasia; almost all aphasic patients produce a diverse range of errors, although the relative proportions of different error-types may differ.

Thus, the differences between normal error studies and aphasic error studies appear to reflect a general loosening of probabilistic constraints—contextual influences, syllable position constraints, markedness constraints, editorial constraints—within the aphasic population (but see Béland et al., 1993), rather than a distortion of linguistic representations, or the operation of abnormal speech production processes. According to Talo (1980), "[a]lthough all kinds of errors occur in both the normal and the pathological corpus, there is a clear difference between the error types in the two groups, in a quantitative sense" (p. 85). Dell (1997b) refers to this phenomenon as a example of the 'continuity thesis', whereby aphasic behaviours can be represented along a quantitative continuum of performance, from normal to severely impaired.

The types of errors observed in the spoken language of normal and aphasic speakers, and the factors constraining those errors, have helped to reveal the psychological structures and processes involved in speech production. Broad similarities between the patterns of normal and aphasic errors have further suggested that the same processing assumptions apply to both populations. Advances in language modeling approaches have helped to show how this is possible. An overview of the major theories of speech production, and how speech error evidence has contributed to their formulation, follows.

Chapter 3. Speech Production Models

Formulating a model of normal speech production requires specification of the linguistic representations involved, how they are combined in the processes of formulating a verbal message, and how those processes might be disrupted. Modeling efforts have benefited from a variety of approaches converging on the study of language: in particular, psychology, linguistics, aphasiology and, most recently, computational modeling. Theories of normal production have been tested by data from aphasia studies, and certain modifications made, for a valid model must be able to account for abnormal as well as normal behaviour (Baars, 1992b; Buckingham, 1980). In turn, the ability of speech production models to account for both normal and aphasic speech errors has contributed to our understanding of the nature of language breakdown in aphasia.

Sentence Production

Although the focus of the present study is the phonological lexicon, a brief review of sentence production models will be presented first, in order to put the phonological processes under investigation into context. Several researchers have relied heavily on speech error evidence in order to articulate models describing how connected speech is produced (e.g. Fromkin, 1971; Garrett, 1975; Levelt, 1989). (For reviews, see Bock, 1995; Bock & Levelt, 1994; Butterworth, 1981; 1992; Cutler, 1995; Fowler, 1995; Fromkin, 1993; 1992; Garrett, 1988; Levelt, 1992; and Meyer, 1992.) They are largely in agreement about the levels of representation and the stages of production, although there is on-going debate about the degree of interactivity of the stages during production, as will be discussed later. A review of some of the most influential of these models illustrates how our understanding of speech production has evolved, as evidence from spontaneous speech production and experimental production paradigms has accumulated.

The Utterance Generator (Fromkin)

As a starting point, Fromkin's Utterance Generator (1971; 1993) outlined six stages: the first three involved the generation of the meaning of the message, its syntactic structure, and its intonational contours; the fourth stage involved lexical selection; the fifth phonological specification; and the sixth generated the motor commands for speech. The types of errors that have been observed (see Table 1-i in Chapter 1) provide evidence that every stage is vulnerable to disruption in some form, but it is the fourth and fifth stages—the retrieval and phonological specification of lexical items—which are of most relevance here. To account for her own observations, Fromkin (1971) speculated that the lexicon must contain a complete list of fully specified formatives, cross-referenced according to semantic class, syntactic category, orthography, and various aspects of phonology, such as number of syllables, rhymes, and shared phonemes.

During speech production, lexical items are retrieved by first consulting the semantic sub-section of the lexicon for items which fit the required semantic features, then following directions to a specified address in the main vocabulary listing. Lexical selection errors occur when the semantic features specify the wrong address (resulting in a semantically related error), or when the correct address leads to a wrong word in the vicinity of the correct word (resulting in a phonologically related error). Once the word form is retrieved, its segments are placed in a buffer in short-term memory, and it is here that segmental misordering errors occur. Although words are retrieved in phonological form at stage four, they are not fully encoded phonetically until after the application of morphophonemic rules in stage five. The retrieval of abstract lexical forms is differentiated from their subsequent morphological and phonological specification in order to account for findings of morphological and phonotactic accommodation, and the separation of stems and affixes in errors. (See also Butterworth, 1992; Garrett, 1988; and Levelt, 1989.) These two stages are also important in making the distinction between aphasic impairments of phonological selection and phonetic implementation (Buckingham, 1987).

Functional and Positional Levels (Garrett)

Garrett's model of sentence production (1975, 1984, 1988) follows a similar sequence, but focuses more on processes than on representations, further specifying the integration of lexical items into syntactic structures. At the 'functional level', semantic-lexical items are selected based on the semantic specifications of the message, assigned functional roles based on the message's syntactic structure, and inserted into the sentence planning frame. Semantic errors occur through the misselection of lexical items, and word ordering errors occur when words are assigned to the wrong slots, but their functional specification accounts for grammatical class constraints in such errors. At the 'positional level', word forms are retrieved, their segments are inserted into syllabic planning frames, and prosodic structure is assigned. The differentiation of frame construction (both at sentence and word levels) from the selection of linguistic components (words and phonemes) to fill the slots in the frame neatly accounts for many of the constraints described earlier. For example, by defining the pool of items being considered for selection, the planning frames constrain potential word substitutions according to their syntactic class, and potential phoneme errors according to their syllabic position.

In the first instantiation of Garrett's model (1975), the positional level was not fully developed, and sound errors were vaguely hypothesized to arise from failures in the assignment of lexical items to their slots. In its more recent version (e.g. Garrett, 1984, described above), however, Garrett incorporates the same sort of frame-filler mechanism that operates at the functional level, which is similar to that described by Shattuck-Hufnagel (1979). Garrett (1988) comments on the apparent redundancy of such a system: "The separation of lexical content from phrasal frame is required by the productivity of syntax... But the same kind of processing separation holds for the separation of segmental content of lexical items, where, seemingly, it is not imperative: the lexical inventory is a finite set" (p. 81). However, he goes on to justify the mechanism as a way to avoid 'excess baggage' early on in the lexical access process when it is not required. That is, the system is hypothesized to operate more efficiently if it has access only to the information which is relevant to each stage; at the stage where lexical items are slotted into their frames, their phonological form is irrelevant, so phonological specification occurs at a later stage.

The Scan-Copier (Shattuck-Hufnagel)

The process by which contextual errors—shifts, exchanges, anticipations and perseverations—occur has been further specified by a 'scan-copier' mechanism (Shattuck-Hufnagel, 1979, 1992). As in Garrett's (1984) model, this frame-content device scans the pool of available items (words or phonemes) for the one which matches the specified constraints (functional or positional), and copies it into the appropriate position. A sequencing error occurs when an item is mistakenly copied into the wrong slot, although (as we have seen) even these errors tend to respect constraints of syllabic position. Shattuck-Hufnagel expands on Garrett's model by including a 'check-off monitor', the function of which is to delete items from the pool once they are assigned a position. Malfunctions of the check-off monitor help to explain how errors might occur: a perseveration occurs when an item is not 'checked off' after use, so it remains available to be inserted again into the utterance, provided there is an appropriate slot (e.g. *black boxes* > *black bloxes**); when an item is copied into an earlier-than-intended slot, and not checked off, an anticipatory error results (e.g. *reading list* > *leading list*). Both these types of errors involve copying or doubling of phonemes. Exchange errors, on the other hand, involve misordering, but because no 'doublets' are created, no breakdown in the check-off monitor has occurred (e.g. *snow flurries* > *flow snurries**). (Examples are from Dell, 1986.) Although non-contextual errors—substitutions, additions, and omissions—might also be explainable using a scan-copier device, the mechanisms of misselection (from outside the pool of items included in the utterance) are not specified in this model.

Evidence of syllabic structure constraints supports the existence of the scan-copier, because it incorporates the hierarchical separation of abstract syllabic frames from the smaller content units which fill the frames (e.g. Stemberger, 1990). The scan-copier has also been adopted to account for aphasic errors (e.g. Buckingham, 1986). The repeated phoneme effect described in the previous chapter, wherein repeated phonemes in the utterance (as in tongue twisters, for example) increase the rate of error production, has been attributed to a hyper-sensitive error monitor (Buckingham, 1987). Buckingham points out, however, that such a monitoring device can also derail in the opposite manner, by not being sensitive enough and thus allowing contextual errors to slip through. For example, the perseveratory nature of neologistic jargon implicates a sequencing deficit such as this (Buckingham, 1990). Furthermore, the sensitivity of the monitor may be affected by factors such as the frequency and complexity of the target utterances (Buckingham, 1987).

From Message Generation to Monitoring (Levelt)

Like Fromkin's and Garrett's models, Levelt's model of speech production (1989) proceeds in sequential stages from message generation to grammatical encoding to phonological encoding to articulation. However, Levelt advanced former models in several important ways, two of which are most relevant here—his conceptualization of the lexicon, and the integration of speech monitoring. Within Levelt's model, lexical representations are divided into two separate components, one for 'lemmas', which represent semantic and syntactic information, and one for phonological word forms, also called 'lexemes'. Lemmas are retrieved during grammatical encoding, which gives rise to the surface structure of the utterance; word forms are then retrieved during phonological encoding, which gives rise to a phonetic plan. Having two separate stores resolves the over-abundance of information that needed to be stored in the unitary lexicon as conceived by Fromkin (1971). This structure also accounts for much of the evidence from speech-error studies, to be discussed in the next section.

Levelt's second major contribution to speech production models was the incorporation of monitoring and editing functions. The ability to monitor utterances and repair errors, not only after they are produced, but also at intermediate stages during their formulation, is an integral aspect of Levelt's model, one which is called upon to explain many of the findings in error research (Levelt, 1983). The experiments of Baars and his colleagues (see Baars, 1992a; and Mattson & Baars, 1992 for reviews) illustrate anticipatory editing at different levels by attempting to elicit errors "designed to meet or violate a large number of semantic and pragmatic criteria" (Baars, 1992b, p. 137). For example, in spoonerism-elicitation tasks, they have found that errors are more likely to occur if they form words as opposed to non-words, syntactically appropriate instead of anomalous phrases, or phrases which do not violate social taboos as opposed to ones that do (Baars, 1992b). The idea that we can monitor speech before it is produced makes intuitive sense, in that we are often aware of incipient errors, and can halt their production. This phenomenon has been investigated experimentally through the generation of tongue-twister errors in inner speech (Dell & Repka, 1992).

Neither of these concepts was unique to Levelt's model. Butterworth (1980), for example, had previously proposed separate semantic and phonological lexicons, and many investigators had previously described editing operations (e.g. Baars et al., 1975; Shattuck-Hufnagel, 1979). Kempen and Huijbers (1983) outlined a process of lexicalization in which an abstract pre-phonological (or L1) item is retrieved, then checked by a monitor for its fit to the requirements of the utterance, before retrieval of the phonological shape (or L2 item) corresponding to the selected L1 item. Despite having many precursors, Levelt's model has been valuable in creating a more complete picture of speech production than had previously been described, and has also engendered a good deal of discussion by virtue of its strict seriality, a point which will be revisited shortly. In its most recent instantiation (Levelt & Wheeldon, 1994; Levelt et al., 1999), Levelt's model also incorporates a syllabary at the level of phonetic encoding (after phonological encoding). This processing component has the advantage of reducing the programming load required by segmental assembly, a problem which both Garrett (1988) and Shattuck-Hufnagel (1992) had identified.

Issues of Lexical Representation and Access

The stage models of speech production outlined above have formed the basis for theories of the storage and retrieval of lexical items during speech production. However, they are not without controversy. The way in which words are retrieved, and what factors influence their successful retrieval, are still the subject of much debate: Is there one lexicon or two? Do they interact in production? Are items in the phonological lexicon abstract or encoded? Another point of contention has been the nature of word-finding in aphasia. Do aphasic speakers retrieve words by the same mechanisms as normal speakers, or do their errors suggest the operation of pathological processes? Psychological theory and methods of experimentation, current linguistic theory, and computational modeling have all contributed to the effort to resolve these issues.

Semantic and Phonological Lexicons

The separation of semantic and syntactic information from phonological information in the lexicon has been proposed by many investigators (e.g. Butterworth, 1980; 1981; Dell & O'Seaghdha, 1992; Dell et al., 1997b; 1993; Fromkin, 1971; Garrett, 1975; 1988; Kempen & Huijbers, 1983; Kohn & Goodglass, 1985; Levelt, 1989; 1992; Roelofs, 1992; Stemberger, 1985; but see Caramazza, 1997). Evidence comes from several sources, error studies being the most obvious. The observation of two distinct types of errors—semantically related errors, assumed to arise at the stage of lemma selection, and phonologically related errors, assumed to arise at the stage of lexeme selection—suggest that these two types of information are retrieved (or not) independently of each other, and thus stored separately. That these errors are indeed 'distinct' has been supported by analyzing the time course of lexical retrieval (Levelt et al., 1991a; 1991b; Schriefers et al., 1990). Using a picture-word interference paradigm, Schriefers and colleagues demonstrated that semantic priming effects were confined to early stimulus-onset-asynchrony (SOA) conditions, during the hypothesized time of lemma access, whereas phonological priming effects only showed up at longer SOAs, during lexeme access. These results were supported by further priming studies (Levelt et al., 1991a). The 'tip-of-the-tongue' phenomenon (e.g. A.S. Brown, 1991; R. Brown &

McNeill, 1966), in which the meaning of the word is known but its form is inaccessible, is further evidence that lemma selection may occur without lexeme retrieval, or with incomplete lexeme retrieval. On the other hand, some evidence from tip-of-the-tongue experiments showing that both grammatical and phonological information may be available has been cited as evidence against the lemma/lexeme distinction (Caramazza & Miozzo, 1997), although this conclusion rests on the assumption that lemma and lexeme access are discrete sequential stages. Although the details are not relevant to the current study, Caramazza's alternative model sets syntactic information apart from phonological and semantic information. Nevertheless, this model also maintains the dual-stage nature of lexical access (Caramazza, 1997; Caramazza & Miozzo, 1997).

In the aphasic literature, this dual-stage hypothesis view has received support from the identification of different types of anomia, characterized by a primary deficit to either the semantic system (e.g. Howard & Orchard-Lisle, 1984) or the phonological system (e.g. Kay & Ellis, 1987). (See also Goodglass et al., 1976.) These can be related to deficits in lemma retrieval and lexeme retrieval, respectively. Such evidence is insufficient to justify the independence of semantic and phonological lexicons, because cases with pure deficits are more rare than cases displaying both types of errors. However, closer analysis of the errors produced by aphasic speakers provides further support. Badecker and colleagues (1995), for example, report a case of anomia in which the Italian patient was able to retrieve grammatical information about a word's gender (encoded in the lemma), but none of its phonological specification. His preserved ability to repeat and read aloud, but not name, gender-marked words localized the deficit to retrieval from the phonological lexicon rather than a post-lexical output stage. Furthermore, the dissociation between grammatical and phonological retrieval was maintained even on exception words, illustrating that access to gender information was not an artifact of inferences made from phonological form. Neurological support for a distinction between semantic and phonemic disorders comes from an analysis of lesion site data in two groups of fluent aphasic subjects (Cappa, Cavallotti, & Vignolo, 1981). Subjects with predominantly phonemic errors (i.e. phonemic paraphasias, *conduites d'approches*, neologisms, and phonemic jargon) presented with lesions close to the sylvian fissure, whereas subjects with predominantly semantic errors (i.e. anomia, circumlocutions, verbal paraphasias, and verbal jargon) had lesions farther from the sylvian fissure.

Thus, the evidence clearly supports separate lexical stores for phonological and semantic (and possibly syntactic) information, but what is the motivation for such an organization? Dell (1997b) points out that "the mapping between concepts and phonological form is a mapping between two unrelated spaces" and that this "arbitrary relation between form and meaning motivates an intermediate step if the mapping is carried out by spreading activation" (p. 804). This is a purely mechanistic constraint, but Dell (1986) further proposes that the separation of different types of information is advantageous for speech production. The productivity of language—that is, the ability to create novel utterances—is allowed by the representation of units of different sizes at different levels, which may therefore be recombined in an infinite variety of ways. However, as mentioned in the previous chapter, this flexibility entails a trade-off in allowing for the creation of errors: "A slip is an unintended novelty" (Dell, 1986, p. 286).

Interactivity vs Modularity

The models of Fromkin, Garrett and Levelt operate in a strictly top-down, serial manner. According to Levelt and colleagues (Levelt et al., 1999; Levelt et al., 1991a), although multiple lemmas may be activated during production, "it would be counterproductive to activate the word forms of all active lemmas that are *not* selected"

(Levelt et al., 1999, p. 15), as they could only interfere with the activation of the selected item. Thus, processing at the semantic level must be complete, and a single candidate selected, before processing at the phonological level begins. One consequence of this is that each stage is "blind with respect to the type of information used by the other stage" (Dell & Reich, 1981, p. 612). However, the independence of the two stages of lexical retrieval has been called into question by findings that semantic and phonological factors can interact in error production. As described in the last chapter, several investigators have found that mixed errors (i.e. those bearing both semantic and phonological relationships to the target) occur at greater than chance levels (e.g. Dell & O'Seaghdha, 1992; Dell & Reich, 1981; Garnsey & Dell, 1984; Harley, 1984). The lexical bias previously described (e.g. Baars et al., 1975; Dell & Reich, 1981) has been cited as further evidence of interactivity between semantic-lexical and phonological levels. Explanations of aphasic deficits also show this conflict between modularity and interactivity, or serial and parallel processing. Deficit patterns have been described specific to retrieval from either the semantic or phonological lexicon, as noted above. Within the phonological system, different deficits have also been described at the stages of lexical (word-form) retrieval and phonemic planning (e.g. Gagnon & Schwartz, 1997; Kohn & Smith, 1994a; 1995). However, many aphasic error patterns resist compartmentalization into a specific level, and have been difficult to explain without hypothesizing multiple deficits.

To explain such findings, lexical access has been proposed to take place by the continuous and overlapping spreading activation from the lexical level to the phoneme level, with reverse feedback connections strengthening related lexical items (Dell, 1985, 1986; 1988; Dell & O'Seaghdha, 1991, 1992; Dell & Reich, 1981; Harley, 1984, 1993a, 1993b; Harley & MacAndrew, 1992). According to Dell and Reich (1981), independence predicts strong constraints on the occurrence of mixed errors, constraints which are not supported by the evidence, whereas interaction predicts the sort of probabilistic influence on error incidence that is shown in speech-error data. This explanation has been criticized as unmotivated within a normal system, because its only function appears to be to account for errors (Levelt et al., 1991b; Nickels & Howard, 1995), whereas modularity serves as "nature's protection against error" (Levelt et al., 1991b, p. 618). However, Dell (1985) proposes that feedback actually helps to filter out potential errors in a normally operating system, by reinforcing the activation of the target. He explains, "positive feedback connections, acting in concert with the primary connections, mold the activation pattern of a lower level until it meshes with information available at higher levels", thus forming a "stable coalition" of activation (p. 5). Furthermore, it has been suggested that feedback connections from phonemic to lexical items may also be used by the comprehension system (Croot, Patterson, & Hodges, 1998; Martin et al., 1994; Martin & Saffran, 1992; see also Fay & Cutler, 1977).

Interactivity is not the only way to account for mixed error and lexical bias effects, however. Within discrete-stage models, findings of interactive effects have been explained by postulating an editing mechanism which discriminates against non-word and non-related errors more than real-word and related errors (e.g. Baars et al., 1975; Butterworth, 1981; Garnsey & Dell, 1984; Kempen & Huijbers, 1983; Levelt, 1989; Levelt, 1983; Nooteboom, 1980). Shattuck-Hufnagel's (1979) scan-copier includes not only a check-off monitor or 'bookkeeper', but also an output error monitor which checks for certain types of sequencing errors. Baars (1980) proposes that the functioning of the editor may be disrupted, and errors thus elicited, by restricting the time available for correcting errors on-line.

Such an editor has also been important in explanations of aphasic deficits. Although Ellis (1985) hypothesized that formal (i.e. real-word, phonologically related) paraphasias would not be expected in a disrupted production system if they rely on the normal operation of monitoring and editing processes, there is evidence that these errors do occur beyond chance levels (Blanken, 1990). This suggests that the editor itself may be disrupted by brain damage. In fact, aphasic speakers have been differentiated by their ability to monitor and correct their own output for errors (Hofmann, 1980; Kohn, 1984; but see Schlenk et al., 1987), and by their ability to adopt different strategies in the face of lexical retrieval deficits (Buckingham & Kertesz, 1974; Liederman, Kohn, Wolf, & Goodglass, 1983).

The use of an editor to explain certain speech errors has been criticized for its *ad hoc* nature (Dell, 1990), although it too may be independently motivated by the need for the comprehension system to monitor incoming speech. However, efforts to draw parallels between production and comprehension deficits in aphasia have indicated that there is no straight-forward relationship between the two (e.g. Miceli, Gainotti, & Caltagirone, 1980; Nickels & Howard, 1995). The answer may lie in a compromise between the two models. Dell and his colleagues (Dell & O'Seaghdha, 1992; Dell et al., 1997b) have proposed a 'reconciliation' of modular and interactive approaches whereby activation spreads interactively throughout the lexicon, but "the selection processes associated with each step are modular" (Dell et al., 1997b, p. 807). Others have suggested that findings from speech error studies mandate at least some degree of parallel, or overlapping, processing across levels, but that complete interactivity is not necessary (Buckingham, 1986; Laubstein, 1999).

Phonological Access and Phonemic Encoding

Beyond lemmas and lexemes, a further distinction has been made between phonological errors which originate at the level of lexeme access, and those which originate at the level of phonological encoding. In normal speakers, these levels of processing differentiate malapropisms from segmental substitution and sequencing errors, a distinction which is not usually difficult to make, given that most segmental errors in normal speech are contextual in nature. In aphasic speakers, the distinction becomes blurred, yet is of greater importance theoretically, in order to identify the nature of the deficit involved. A number of studies have attempted to find evidence for these two distinct sources of error by analyzing the phonological characteristics of word and non-word errors. In one study, a large corpus of neologisms produced by fluent aphasic subjects was analyzed for bimodality in the degree of target relatedness; none was found, suggesting that all the errors arose from the same source (Gagnon & Schwartz, 1996). In another study, phonemic paraphasias produced by fluent aphasics were compared to a pseudo-corpus to look for a lexical bias; a greater than chance rate of word production among the paraphasias suggested a lexical source for at least some of these errors (Gagnon et al., 1997). Kohn and Smith (1994a) compared the patterns of phonological errors shown by two different subjects, analyzing such variables as the proportion of formal paraphasias, the degree of phonological relatedness between errors and targets, and the serial positions of segmental errors. Results suggested that the two subjects, one with Wernicke's aphasia and one with conduction aphasia, showed two distinct functional deficits—the first in lexical-phonological activation and the second in phonemic planning.

Determining the source of neologisms has proven to be a particularly thorny problem in aphasia research, because it is often unclear, even in structured tasks like picture naming, what the intended target of the utterance is. Several hypotheses about neologism production have been advanced, based on the phonological characteristics of the utterance, the degree to which a target is identifiable, and the way in which the pattern of errors changes as the aphasia resolves. Originally, it was proposed that abstruse neologisms were simply phonemic paraphasias which were distorted beyond recognition (Buckingham, 1977; Kertesz & Benson, 1970; Lecours & Lhermitte, 1969). Kertesz called this a 'conduction defect' because it was supposed to be due to an "excessive accumulation of literal paraphasias" (p. 385), characteristic of conduction aphasia. This view was later challenged on the basis of observations that neologisms usually appear to replace nouns, that neologisms often co-occur with word-finding difficulties and with perseveration, and that neologistic jargon tends to resolve to anomia, rather than a pattern of phonemic paraphasias. These findings suggested an underlying lexical access deficit, rather than a phonological encoding deficit (Buckingham, 1977, 1987).

An alternative hypothesis has been put forth, that neologisms come from semantic paraphasias which are subsequently phonologically distorted, a so-called 'two-stage error'; (Pick, 1931, cited in Buckingham, 1981; Butterworth, 1992; Howard, Patterson, Franklin, Orchard-Lisle, & Morton, 1985; Lecours & Rouillon, 1976). For many neologistic errors, this explanation makes intuitive sense, if the intervening semantic substitution is discernible, for example: *web* > /spaidɪd/ (Howard et al., 1985). However, such a connection is often difficult to judge unambiguously (Buckingham, 1990; Butterworth, 1992). Furthermore, this mechanism requires two independent sequential errors and, thus, may not be the most parsimonious explanation for the majority of neologisms.

For those neologisms without an identifiable relationship to a target or to other words in the context (except perhaps to other neologisms), Butterworth (1979, 1992) proposed that aphasic speakers (albeit unconsciously) resort to the use of a compensatory device. This device "quasirandomly combines English phonemes in a phonotactically regular way" (Butterworth, 1979, p. 133); thus it has been labelled the 'random generator' (e.g. Buckingham, 1981; 1990b). Evidence for such a device has come from Butterworth's original study (1979), which showed that hesitations preceding non-target-related neologisms were longer than those preceding target-related neologisms, suggesting the operation of a time-consuming mechanism. Findings that abstruse neologisms obey phonotactic constraints (Butterworth, 1979) and tend to include higher-frequency phonemes; (Code, 1982; but see Butterworth, 1979), and the observation that neologistic production tends to resolve to characteristic anomic symptoms (Buckingham, 1977, 1981; Green, 1969) support the operation of a random generator as a compensatory device.

Nonetheless, the concept has been criticized for its dependence on a mechanism presumed to be created anew following brain injury, an unlikely turn of events (Ellis et al., 1983). However, it has been justified as a default mechanism, along the same lines as the constraints which involve the substitution of unmarked for marked syllable structures in normal phonological errors (Butterworth, 1992). In fact, Buckingham (1987) likened the random generator device to Sussman's (1984) model of a neuronal syllabic template, in that both involve the production of "strings of phonotactically acceptable syllables, dissociated from the lexical inventory of the language" (p. 387), and cited as supporting evidence the ability of some speakers to produce similar-sounding 'voluntary glossolalia' (Lecours, 1982). Buckingham (1990b) characterized the random generator as "an alternative way of describing a system of phonological knowledge that all speakers possess as part of their cognitive linguistic machinery...a normal, albeit underused, capacity" (p. 215).

Alternative Paradigms

Modeling of the psychological processes involved in speech production has benefited enormously from the adoption of theories and methods from other fields. In particular, the incorporation of insights from modern linguistic theory, and the use of computational techniques to simulate language production have complemented the psychological models outlined, and informed our understanding of how speech errors occur.

Current Linguistic Theory

The importance of linguistic knowledge to any explanation of pathological language breakdown has long been recognized (Bierwisch, 1981; Blumstein, 1973b; 1990; Jakobson, 1964; MacMahon, 1971). Lecours (1990) asserted that once Jakobson had set the stage, "achieving a certain level of complementarity between structural linguistic characterizations and brain-compatible psycholinguistic models has since become a more or less explicit objective of aphasiology" (p. 116). Béland (1993) was not so optimistic, claiming that "neuropsychologists consider that accessed phonological representations are very close to their surface form and tend to minimize the amount of processing involved in speech sound production" (p. 284).

Nevertheless, many recent studies, including Béland's own, have used linguistic principles such as sonority, markedness, and consonant harmony to help reveal the mechanisms underlying aphasic speech errors (Bastiaanse et al., 1994; Blumstein, 1973b; Christman, 1994; Favreau et al., 1990; Kohn et al., 1998; Kohn et al., 1995; Nespoulous et al., 1984; Romani & Calabrese, 1998). Wheeler and Touretsky (1997) offer an account of normal and aphasic contextual speech errors that relies on parallel licensing constraints: underspecification at underlying levels allows for multiple licensing of segments, which nevertheless obey phonotactic constraints; errors arise from the incorrect resolution of these licensing conflicts. The authors claim that their model provides a unitary account of normal and aphasic errors, without relying on special 'error-generating' devices like Shattuck-Hufnagel's (1979) scan-copier and the random generator (Buckingham, 1990). According to Béland (1993), whereas earlier studies (e.g. Blumstein, 1973a) treated phonological errors as transformations of one surface representation to another, studies taking current linguistic theory into account reveal that "errors can arise at different levels and... result from the application of universal phonological processes" (Béland et al., 1990, p. 159).

Connectionism and Computational Modeling

Connectionist approaches take an entirely different perspective from theoretical linguistic approaches, representing "a shift of emphasis from symbolic, rule-based processing to sub-symbolic systems that compute their outputs from the interaction of many simple, interconnected neuron-like units" (Harley, 1993a, p. 221). According to Harley (1993a), connectionist models have a number of advantages: for example, being patterned on the neuronal interconnectivity of the brain, they have 'biological plausibility'; they allow 'graceful degradation' of function rather than the all-or-nothing disruption implied in many modular and rule-based approaches; they are able to satisfy multiple constraints simultaneously without requiring the explicit encoding of rules; they are able to illustrate how normal and disordered systems are related, and how the same patterns of behaviour can arise from different lesions. This last point is also a disadvantage, because it means that a given pattern of errors may be equally well explained by different models (e.g. Wright & Ahmad, 1997). To some extent this is unavoidable; as Stemberger (1985) notes: "The data available... always underdetermines theory; more than one theoretical description is always compatible with the data" (p. 10). Another criticism of connectionism is that the ability to alter a given model's parameters infinitely makes it very difficult to falsify (e.g. Nickels & Howard, 1995). The same criticism, however, can and has been levelled at modular theories, for their proliferation of boxes and arrows. (Harley points out, for example, that "it is difficult to envisage what data could either falsify or verify the editor model" (1984, p. 212).)

In connectionist network models, errors arise from an unintended element reaching a higher level of activation at the moment of item selection. This happens when there is 'noise' in the system (Dell & O'Seaghdha, 1991), arising from either random fluctuations in resting levels of activation, variation in relative activation levels due to different and changing frequencies of use, or the spread of activation from non-target units (Dell, 1986; Dell & Reich, 1981; Stemberger, 1985). Spread of activation may result in semantically related errors, such as the blending of synonyms, or the substitution of antonyms or category members, whose lemma nodes are connected within the semantic network (Roelofs, 1992). Phonologically related errors are produced through activation spreading from lexeme nodes to their constituent phoneme nodes and rebounding back to the lexeme network. The role of noise in the system also helps to explain the inconsistency of performance in aphasia (Harley & MacAndrew, 1992), because relative activation levels may change as aspects of the linguistic and non-linguistic context change.

Connectionist modeling has been facilitated by the application of computer simulations, which may be programmed to match data from normal speakers, then 'lesioned' for comparison to aphasic speech production (e.g. Dell et al., 1997b; Harley &

MacAndrew, 1992; Rapp & Goldrick, 2000; Wheeler & Touretzky, 1997). Aphasic error patterns have been successfully simulated by modifying such parameters as strength of connections among nodes (e.g. Croot et al., 1998; Lavorel, 1982; Schwartz et al., 1994), the activation-to-noise ratio (e.g. Laine & Martin, 1996; Schwartz & Brecher, 2000; Wright & Ahmad, 1997), or the rate at which activation decays (e.g. Martin et al., 1994; Martin et al., 1998; Martin & Saffran, 1992). Harley and McAndrew (1992) compared these three types of lesion and found that a reduced flow of activation best accounted for the data from aphasic subjects. Dell and colleagues (1997b) found that different aphasic error patterns could be simulated using different types of 'lesions': decay lesions promoted more 'normal' errors such as semantic and mixed word substitutions; connection-weight lesions promoted more severe distortions of language, like non-word errors and unrelated word substitutions. It has been noted, however, that not all aphasic error patterns can be well fit to a weight/decay model (Nickels & Howard, 1995; Schwartz & Brecher, 2000). Because deficits can be characterized by such global processing impairments, these modeling efforts have provided strong support for interactivity in speech production (Rapp & Goldrick, 2000).

Investigations using both linguistic analyses and computational modeling have illustrated parallels between normal and aphasic speech production—linguistics by illustrating that many pathological surface structures actually reveal the operation (or hyper-operation) of normal constraints (Béland et al., 1990); computational modeling by illustrating the variety of outcomes that can arise from global quantitative changes to a number of parameters (Dell et al., 1997b). In many ways the two are also complementary. For example, linguistic theory has been particularly valuable in accounting for contextual errors (e.g. Wheeler & Touretzky, 1997), whereas connectionist models provide a framework by which non-contextual substitutions can be explained (Dell et al., 1993; Dell et al., 1997b). So far, however, connectionist models have been limited in their scope to modeling single-word lexical access using restricted vocabularies (Dell et al., 1997b; Harley, 1993a). In order to provide not only a description of how phonological errors come about, but also a motivation for why they might occur, it is necessary to specify the architecture of the phonological lexicon, and to provide a motivation for its structure. The following section describes efforts which have been made to address this problem.

Similarity Neighbourhoods

Although the structure of the phonological lexicon has not been operationally defined until recently, phonological neighbourhoods have been implicitly assumed in a number of different experimental paradigms. Following demonstrations of semantic priming (e.g. Collins & Loftus, 1975), phonologically related words have also been shown to affect the recognition of written words (e.g. Columbo, 1986; Hillinger, 1980; Meyer, Schvaneveldt, & Ruddy, 1974) and spoken words (e.g. Slowiaczek & Hamburger, 1992; Slowiaczek, Nusbaum, & Pisoni, 1987). In auditory word recognition, word onsets appear to be especially important (Grosjean, 1980; Marslen-Wilson & Zwitserlood, 1989; Marslen-Wilson & Welsh, 1978; Salasoo & Pisoni, 1985; but see Luce, 1986; Slowiaczek & Pisoni, 1986), a finding that is not surprising, given the temporal nature of auditory word recognition. To explain this finding, Marslen-Wilson and Welsh (1978) developed the 'cohort model' of auditory word recognition, in which a set of words matching the incoming stimulus is activated and gradually narrowed down as more of the auditory signal becomes available. Thus, Marslen-Wilson and Welsh defined one kind of 'similarity neighbourhood' based on shared word onsets. Somewhat counter-intuitively, rhyme relationships have also been shown to have facilitative effects on auditory word recognition (Baum, 1997; Burton, 1989;

Gordon & Baum, 1994; Milberg, Blumstein, & Dworetsky, 1988; Slowiaczek et al., 1987), suggesting that a word's neighbourhood needs to be more broadly defined than its word-initial cohort. More recent instantiations of cohort theory also recognize that non-initial shared phonology may play a role in finding a 'best-fit' match of an auditory stimulus to a target (e.g. Marslen-Wilson, Moss, & Van Halen, 1996; Marslen-Wilson & Zwitserlood, 1989). For aphasic subjects, the effectiveness of phonemic cues—especially initial phoneme cues, but also rhymes—is further evidence of the importance of phonological relationships in lexical access (Pease & Goodglass, 1978; Spencer et al., 2000).

In speech recognition, relationships among phonologically related words are hypothesized to be important in allowing the perceptual system to overcome noise and ambiguity in the acoustic input. It is apparent, however, that phonological relationships among words also affect output processes. During tip-of-the-tongue states, speakers frequently have access to the initial phoneme of the missing word (e.g. A.S. Brown, 1991; R. Brown & McNeill, 1966), providing support for a 'cohort-like' concept of phonological neighbourhood in production as well as perception. Other information is also often available, such as the target word's length and stress pattern, which is difficult to explain without a more holistic view of the phonological lexicon. Phonologically related speech errors also provide compelling evidence that a wide variety of types and degrees of target-error relatedness affects speech production. Thus, there is an abundance of evidence for a phonologically organized lexicon, but the nature of its organization remains unspecified.

Landauer and Streeter (1973) pointed out that well-known effects of frequency in word recognition may be influenced by differences between common and rare words which had hitherto been overlooked. They illustrated that common words have many more orthographic 'neighbours', that is words which contain all but one of the same letters. Experimenters began to incorporate measures of neighbourhood size, or density, in their studies of lexical access (e.g. Coltheart, Davelaar, Jonasson, & Besner, 1977). Furthermore, it has since been suggested that the frequencies of these neighbours, relative to the frequency of the target word, may also influence the rate and accuracy of lexical access (e.g. Grainger, 1990). These factors of neighbourhood density and neighbourhood frequency have been manipulated in word recognition experiments to assess their respective roles in lexical access. Results have been complex, sometimes showing facilitative effects, sometimes inhibitory effects, and sometimes null effects, and frequently showing interactions among the three variables of target frequency, neighbourhood density, and neighbourhood frequency (e.g. Andrews, 1989; Grainger, 1990; Sears, Hino, & Lupker, 1995; Segui & Grainger, 1990). To some extent, conflicting findings can be attributed to task differences, such as naming *vs* lexical decision (e.g. Grainger, 1990), the presence *vs* absence of perceptual masking (e.g. Forster & Davis, 1991), or the characteristics of the stimuli (e.g. Forster & Taft, 1994). It has been proposed, however, that effects in opposing directions may be accommodated within the same model because of the potential counter-action of inhibitory lateral influences and facilitative reciprocal (interactive) influences (Sears et al., 1995). The studies described thus far all used written stimuli, and an orthographic definition of neighbourhood (Coltheart et al., 1977), but they have paved the way for similar studies in spoken word recognition and production.

The Neighborhood Activation Model

The Neighborhood Activation Model (NAM, Luce, 1986; Luce & Pisoni, 1998) has introduced an empirically testable model of neighbourhood effects in the study of spoken word recognition and production. In these studies, rather than using Coltheart et al.'s (1977) similarity metric based on orthographic form, neighbours were defined as all those words which differ from the target by one phoneme, either substituted, added, or omitted. Based on the general assumption that word recognition involves a process of discrimination among competing similar lexical items (e.g. Marslen-Wilson & Welsh, 1978), it was hypothesized that word recognition would be slower and less accurate for lower frequency words, for words from more dense or confusable neighbourhoods, and for words with higher frequency neighbours. Results from a variety of experimental paradigms—perceptual identification in noise, auditory lexical decision, auditory word naming (i.e. repetition), and auditory-word identification—have, for the most part, upheld these predictions (Goldinger, Luce, & Pisoni, 1989, 1992; Luce & Pisoni, 1998; Luce, Pisoni, & Goldinger, 1990). These studies illustrated that frequency of occurrence should be considered a relative, rather than an absolute, characteristic; that is, the frequency of a given stimulus word must be considered in comparison to the frequencies of its neighbours. Furthermore, they showed that neighbourhood density exerts an effect on word recognition independent of frequency. To account for these findings, Luce and Pisoni (1998) proposed the Neighborhood Activation Model, which incorporates the frequency of occurrence of a given word with the number and frequency of its phonologically related neighbours in a 'frequency-weighted neighborhood probability rule', which is used to predict the probability that the word will be correctly recognized.

Neighbourhood effects in word recognition have not been entirely consistent, however. For example, in their auditory lexical decision task, Luce and Pisoni (1998) found the expected density effect in reaction time, whereby responses to high-density stimuli were slower than to low-density stimuli, but the accuracy data reflected opposing density effects for word and non-word stimuli. Non-words from high-density neighbourhoods were less accurately identified than non-words from low-density neighbourhoods, but words were identified *more* accurately if they came from high-density rather than low-density neighbourhoods. Furthermore, these density effects were noted for low-frequency but not high-frequency words, and for non-words from high-frequency neighbourhoods, but not low-frequency neighbourhoods. The apparent facilitative effect of density for low-frequency word stimuli—contrary to expectations—was explained as a consequence of the time-limited response required by the task. Because lexical decisions take longer for low-frequency words, decisions forced at the response deadline may be made on the basis of the overall level of activation in the neighbourhood; high-density neighbourhoods, having a higher overall level of activation, are more likely than low-density neighbourhoods to result in a 'word' response.

Nevertheless, other facilitative effects of lexical density have also been found in auditory perception experiments. In phoneme identification experiments, frequency-weighted neighbourhood density has been shown to influence the recognition of non-word stimuli in the same way as lexical status does; that is, to shift the category boundary towards the end of the continuum representing the higher-density neighbourhood (Newman, Sawusch, & Luce, 1999; Newman, Sawusch, & Luce, 1997). Similar results have also been shown for fluent and non-fluent aphasic subjects (Boyczuk & Baum, 1999). In addition, Vitevitch and colleagues (Vitevitch, Luce, Charles-Luce, & Kemmerer, 1997) found that, in 'auditory naming' (i.e. repetition), reaction times were faster for non-words composed of high-probability than low-probability phonotactic patterns. Phonotactic probability refers to the frequency of phonemes and sequences of phonemes (Trask, 1996, cited in Vitevitch & Luce, 1998), and is closely related to neighbourhood density: "High-probability phonotactic patterns are high in probability precisely because there are many words sharing the component segments" (Vitevitch & Luce, 1998, p. 325). Thus, these results conflicted with the predictions of the Neighborhood Activation Model.

Vitevitch and Luce (1998) proposed that facilitative effects of phonotactics and neighbourhood density may be reconciled with the Neighborhood Activation Model by attributing facilitative and competitive findings to different levels of processing. Whereas neighbourhood density exerts a competitive effect at the lexical level, phonotactic probability exerts a facilitative effect at the sub-lexical level. Which of these opposing influences is found depends on the lexical status of the stimuli. This hypothesis was tested by Vitevitch and Luce (1998) in an auditory naming task using real-word and non-word stimuli divided into groups of high and low density and high and low phonotactic probability values (because density and phonotactic probability are confounded, the two variables could not be independently manipulated). The stimulus sets were equated on frequency. As predicted, opposite effects of density were found for word and non-word stimuli: low-density/low-probability (LD/LP) words were repeated *faster* than high-density/high probability (HD/HP) words, showing a competitive effect of neighbourhood density, whereas LD/LP non-words were repeated *slower* than HD/HP non-words, showing a facilitative effect of phonotactic probability.

Furthermore, Vitevitch and Luce (1999) illustrated that the level of processing can be manipulated by the requirements of the task. In an auditory lexical decision task, which requires lexical-level processing even for non-word stimuli, inhibitory effects of density were found for both words and non-words. Although sub-lexical processing may also be involved for both word and non-word stimuli, it appears that lexical effects, when they are present, dominate sub-lexical effects (Vitevitch, Luce, Pisoni, & Auer, 1999). Conversely, a same-different matching task, which encourages sub-lexical processing, showed the expected facilitative (phonotactic probability) effect for non-words and, although the inhibitory (density) effect was not reversed for word stimuli, it was no longer significant (Vitevitch & Luce, 1999).

Neighbourhood Effects in Speech Production

More recently, some researchers have also begun to investigate the effects of the phonological neighbourhood on the accuracy of speech production (e.g. Vitevitch, 1997; Vitevitch, ms in prep). Because similar word frequency effects had been found in input and output tasks, it was initially hypothesized that neighbourhood characteristics might also influence production in the same way as recognition (Vitevitch, 1997). Alternatively, it has been suggested that the opposite pattern might be expected, given that word recognition and word production proceed in opposite directions (Best, 1995; Vitevitch, 1997). Nevertheless, an influence of neighbourhood characteristics on lexical production was predicted.

In Vitevitch's (1997) study, targets of malapropisms from the study by Fay and Cutler (1977) were divided into high- and low-frequency, and high- and low-density groups, and compared using a chi-square analysis. Results showed an interaction of target frequency and neighbourhood density, such that, for low-frequency words there were more errors in sparse than dense neighbourhoods, whereas for high-frequency words there were more errors in dense than sparse neighbourhoods. In addition, there were more errors overall for target words from low-frequency than high-frequency neighbourhoods. To compare the neighbourhood characteristics of the malapropisms to the general properties of the lexicon, an equal number of content words was randomly selected from the lexical database to serve as a control word corpus. The control words were found to be significantly different from the corpus of malapropisms and their targets on all three variables: the malapropism corpus was lower in frequency, and had lower neighbourhood density and neighbourhood frequency values than the control corpus. This comparison makes it difficult, however, to determine whether the differences between the corpora are due to the characteristics of the targets or the characteristics of the errors.

In addition to assessing the influence of the phonological neighbourhood on the susceptibility of items to error, Vitevitch (1997) also analyzed the effects of neighbourhood variables on error outcome. Comparing individual error/target pairs showed a slight tendency for errors to be relatively higher in frequency than their targets (54%). However, an ANOVA showed no significant difference overall between errors and targets, on any of the measures of frequency, density, or neighbourhood frequency. Thus, although the role of the phonological neighbourhood on error outcome appears to be negligible, facilitative effects are shown on target susceptibility.

Evidently, the influence of neighbourhood characteristics on speech production stands in sharp contrast to their influence in speech recognition. Rather than showing inhibitory effects of neighbourhood density and neighbourhood frequency, it appears that lexical items benefit from an accumulation of activation spreading from phonologically related items, and are thus less susceptible to error. (An exception is the competitive effect of neighbourhood density shown for high-frequency words. Although Vitevitch (1997) does not attempt to explain this anomalous result, it may be that high-frequency words do not benefit significantly from an increase in neighbourhood activation because they already have a high resting level of activation. It is unclear, however, why an increase in neighbourhood density would reduce speech production accuracy.) The facilitative effects of density and neighbourhood frequency cannot be accounted for by the Neighborhood Activation Model at present. However, Vitevitch suggests that, as in the speech recognition literature, incorporating a sub-lexical level into the model might provide the necessary mechanism to explain such effects.

In order to extend these findings to errors from more controlled tasks, Vitevitch (ms in prep) elicited phonological speech errors using three different techniques. In the first, sound exchanges, or spoonerisms, were induced using the SLIPs technique developed by Baars and Motley (e.g. Baars et al., 1975; Motley & Baars, 1975) (described in the previous chapter). CVC words were combined to create sets of stimuli differing along three dimensions—high- and low-frequency words, words from dense and sparse neighbourhoods, and words with high and low neighbourhood frequencies—providing eight stimulus conditions. Of the errors elicited, only initial sound exchanges were counted. A three-way ANOVA showed that significantly more errors occurred on low-frequency than high-frequency stimuli and on stimuli from sparse than dense neighbourhoods. There was no significant effect of neighbourhood frequency, nor were any of the interactions significant. To explore whether the facilitative effect of density was sub-lexical in origin, analogous to results from recognition experiments, the mean numbers of errors in each condition were correlated with the mean sum of segment frequencies. The correlation was not significant, suggesting that the density effect was not solely attributable to the phoneme frequencies of the stimuli.

A second experiment was conducted to corroborate the density effect. This experiment made use of a tongue-twister task employed in previous studies (e.g. Shattuck-Hufnagel, 1992) in which a string of four similar-sounding words were read aloud six times as quickly as possible. Half of the tongue-twisters contained words from dense neighbourhoods, and half from sparse neighbourhoods, but the two sets of stimuli were statistically equivalent in frequency and neighbourhood frequency. Initial consonants were also controlled across the two conditions, in order to factor out potential effects of phoneme frequency (again only initial consonant errors were counted). More errors were elicited from tongue-twisters with sparse than dense neighbourhoods, supporting the density effect from the first experiment.

In the third experiment, Vitevitch used a picture-naming task in which half of the stimuli were from high-density neighbourhoods, and half were from low-density neighbourhoods. Phonotactic probability, frequency, and neighbourhood frequency were equated for the two sets of stimuli. The latency of each naming response was measured in addition to its accuracy. A main effect of density was shown in the reaction time measure, but not in error rates. This finding was noted by Vitevitch to conflict with a non-significant correlation found in a previous study between latency of naming responses and density (Jescheniak & Levelt, 1994). However, the analysis in the latter study was performed *post-hoc*, and may reflect an insufficient range of densities in the stimuli (Vitevitch, ms in prep).

From the results of these three experiments, Vitevitch (ms in prep) concluded that the facilitative effect of density in production cannot be explained by sub-lexical influences alone, but is due to the interaction of words and phonemes. He interprets his results with reference to the contrasting predictions of a feed-forward model (WEAVER++, Levelt et al., 1999) and an interactive model (Node Structure Theory, MacKay, 1987). Whereas WEAVER++ adopts the same principle of competition among lexical candidates that is used in the Neighborhood Activation Model, Node Structure Theory predicts a facilitative effect of density as a function of the more frequent, and thus more efficient, transmission of activation among word and phoneme nodes in a dense neighbourhood. The strong facilitative effects of density found by

Vitevitch (ms in prep), independent of the effects of frequency and phonotactic probability, clearly support an interactive model.

The facilitative effect of density in speech production has also been supported by studies of the tip-of-the-tongue (TOT) phenomenon (Harley & Bown, 1998; see also Vitevitch & Sommers, ms in prep). According to the 'incomplete activation' hypothesis (e.g. A.S. Brown, 1991; Meyer & Bock, 1992), TOT states occur when a lexical item to be retrieved is only partially activated, resulting in the availability of the concept and perhaps some form-related information, such as the first phoneme or the number of syllables. As a consequence of this partial activation, "some of [the target's] relatives [or neighbours] may be retrieved in its place" (Meyer & Bock, 1992, p. 715). By contrast, the 'interference', or 'blocking', hypothesis (e.g. Jones, 1989), proposes that activated neighbours, or 'interlopers', actually block successful access to the target.

Harley and Bown (1998) addressed this controversy by comparing the incidence of TOT states produced in response to definitions for four different types of words: high-frequency words in high-density neighbourhoods, high-frequency words in low-density neighbourhoods, low-frequency words in high-density neighbourhoods, and low-frequency words in low-density neighbourhoods. Significant main effects were shown for both frequency and neighbourhood density, with more TOTs produced on low- than high-frequency words, and more TOTs on words from low- than high-density neighbourhoods. There was also a significant interaction between the two factors which was not explored, but appears to be due to a larger density effect (albeit in the same direction) for low-frequency than high-frequency words. (This lends support to the idea that the similar interaction found by Vitevitch (1997) for malapropisms was related simply to low-frequency words being able to benefit more than high-frequency words from the facilitative effects of density.) A regression analysis showed a positive correlation between the number of TOTs produced and the length of the target word, so a second experiment was run with only one- and two-syllable targets. The same results were shown.

On trials where interlopers (i.e. errors) were produced, the interlopers were unexpectedly less frequent, with lower neighbourhood densities than their targets. However, the majority (63%) of interlopers were semantically, not phonologically, related. Furthermore, interlopers frequently violated syntactic category constraints. The authors concluded that the processes involved in prolonged, volitional lexical searching are probably not typical of normal automatic lexical access processes (Harley & Bown, 1998).

These studies of normal errors make it clear that the concept of the phonological neighbourhood plays an important role in accounting for the susceptibility of words to error. As suggested more than twenty-five years ago by Landauer and Streeter (1973), the difference between common and rare words runs deeper than their frequency counts. Factors such as length, phoneme frequency, and density, which are confounded with frequency, need to be explored to determine their independent influences on speech recognition and production. These effects may help to explain the inconsistencies noted in frequency effects in speech-error studies. Not only are neighbourhood variables clearly important in speech production processes, they also have distinctly different effects on output than input processes. Whereas phonological neighbours compete with target words during word recognition, they apparently provide activation which reinforces the target during speech production. Adopting a term used by Taraban and McClelland (1987) to refer to the influence of orthographic neighbours on a target's pronunciation, Vitevitch (1997) described the facilitative effects of neighbourhood activation as 'conspiracies' (also called 'gang effects' (Best, 1995)).

In studies of aphasic errors, neighbourhood variables are also starting to be considered. For example, Best (1995) found a 'reverse length effect' in the naming responses of one aphasic subject, such that longer words were easier to name than shorter words. She proposed that, because longer words have fewer neighbours, there is less competition for their access. A *post-hoc* analysis was performed to assess the effect of density (called 'nness' by Best) on a restricted set of two-syllable targets only. Although statistically non-significant, the direction of the nness effect unexpectedly showed *greater* accuracy for items with more neighbours. Thus, even for a subject who showed an atypical effect of length on naming performance, a facilitative effect of density was indicated once length was controlled. Nickels and Howard (1995) also appealed to possible density effects to explain the relationship between lexicality and length in phonological naming errors. Whereas the proportion of non-word errors in their corpus was positively correlated with target length, the proportion of real-word errors was negatively correlated with target length. Given that errors are either words or non-words, these two findings are, of course, interdependent. Nevertheless, an explanation for this trade-off can be found in the characteristics of the lexicon: because shorter words have more neighbours, then phonologically related word errors should be expected to occur by chance more often on short than long words (Nickels & Howard, 1995). The concept of the density of 'lexical space' has also been utilized to estimate chance levels of word outcomes (e.g. Dell & Reich, 1981; Dell et al., 1997b; Gagnon et al., 1997). Notwithstanding these preliminary suggestions and speculations, neighbourhood effects have not yet been explicitly investigated with aphasic subjects.

Chapter Summary

Although speech production models have generally lagged behind models of speech recognition (Cutler, 1995; Fromkin, 1993), theoretical perspectives and modeling techniques from a number of different domains have promoted considerable advances in our understanding of speech production in recent years. The incorporation of two very different approaches—current linguistic theory and computational modeling—into psychological models of language production has allowed further specification of the levels of representation underlying an utterance to be produced, and the ways in which they relate to each other during speech production. In particular, our knowledge of how the lexicon is structured has been advanced through the consideration of sub-lexical structures such as syllabic constituents, which are independently motivated by linguistic theory, and through the power afforded by computational modelling techniques.

To a large extent, models of lexical structure and processing are complementary rather than contradictory (Bock & Levelt, 1994; Levelt, 1992; Nickels, 1997), although there remain areas of controversy and uncertainty. Most researchers agree that the mental lexicon is composed of two separate stores of words—one with semantic features specified, and one with phonological form specified. There is less agreement, however, about the respective roles played by the two lexicons as speech is produced; specifically, are they accessed in sequence, or do they interact? Speech errors have provided strong evidence for the interaction of semantic and phonological information during speech production, but the issue remains open to debate.

Another issue which has only recently been addressed experimentally concerns the structure of the phonological lexicon. Although the idea that words are connected by sound similarity has been a long-standing assumption of research into speech recognition, the implications for speech production have rarely been explored. How might sound similarity facilitate lexical access in production? The Neighborhood Activation Model (Luce & Pisoni, 1998), developed to explore the role of phonological variables in speech recognition, accounts for the influence on a target's recognition, not simply of the *presence* of phonologically related competitors, but also of the *number* of phonologically related competitors, and their frequencies of occurrence relative to the frequency of the target itself (Luce & Pisoni, 1998; Vitevitch et al., 1999). These factors, called neighbourhood density and neighbourhood frequency, have since been shown to influence the accuracy of speech production as well, although in a facilitative rather than a competitive manner (Harley & Bown, 1998; Vitevitch, 1997; Vitevitch, ms in prep). There have been suggestions that neighbourhood variables influence aphasic speech production as well (e.g. Best, 1995; Nickels & Howard, 1995), but the question has yet to be addressed experimentally. The present study was designed to achieve this goal.

Overview of the Present Study

In the past century, from the qualitative descriptions of Meringer and Mayer (1895) and the psychiatric speculations of Freud (1901) to the computational analyses of many current investigators (e.g. Dell et al., 1997b; Harley & MacAndrew, 1992; Rapp & Goldrick, 2000), the study of speech errors has become increasingly sophisticated, both technologically and theoretically. In Chapter 1, the methodology of error study over the years, especially the last twenty-five years, was reviewed. Despite numerous threats to the validity and reliability of error research, the use of a variety of methods, from spontaneous speech studies to experimental elicitation techniques, and the study of errors from both normal and aphasic speakers has provided convergent sources of evidence about the mechanisms and constraints governing error production. This review illustrates the importance of several methodological factors which are taken into account in the present study: 1) the use of different tasks to provide convergent sources of evidence; 2) the collection of enough errors to be considered a representative corpus; 3) the inclusion of subjects with a range of aphasic sub-types to allow greater generalization of the results; and 4) the establishment of a reliable transcription of the error corpus.

Chapter 2 reviewed the research concerning factors which constrain speech errors, focusing on phonological speech errors. The susceptibility of words to error in normal speakers has been shown to be influenced by a number of factors, such as frequency of occurrence, syllabic structure, and linguistic context. Contextually influenced error outcomes are also strongly constrained by the characteristics of the intrusion, as noted, for example, in the high probability that error and target words will be from the same grammatical class, or that error and target phonemes will occupy the same syllable position. Non-contextual errors, however, appear to be less constrained, at least insofar as the constraints can be discerned. Broad similarities are noted between normal and aphasic speech errors, although there are quantitative differences in error patterns. One such difference is in the relative proportions of contextual and non-contextual errors; because aphasic speakers produce more non-contextual errors, it is of theoretical interest to explore the phonological factors which might influence their occurrence. For this reason, the present study examines both contextual and non-contextual errors. Because the constraints studied to date have not been able to adequately account for aphasic error patterns, this study addresses the role of phonological neighbourhood variables, which have not yet been examined in aphasic speech production.

In Chapter 3, a number of speech production models were discussed, in particular those which describe in detail the processes of lexical access. Theories of speech production have been informed by the study of speech errors and they have, in turn, provided powerful explanatory tools of the mechanisms underlying speech production. Nevertheless, there remain large gaps in our knowledge of error production in aphasia. What do aphasic errors reveal about the nature of the connections in the lexicon? What types of phonological relationships are specified, and how might they be represented? To what extent might an understanding of lexical structure and process help to predict aphasic error production and to address such deficits in therapy? The current study investigates, in particular, the structure of the phonological lexicon, the question of interactivity in lexical access for production, and the ability of normal models to account for aphasic patterns.

The aim of the present study is to investigate the nature of phonological speech errors in aphasia, in order to learn more about the processes of normal language production and how it can be disrupted in aphasia. Errors from both spontaneous speech and experimental tasks are analyzed descriptively and quantitatively, to explore the role of the phonological neighbourhood, that is, those items which are considered to be phonological competitors for selection, in contributing to the susceptibility of targets to error and to the nature of the errors produced. Neighbourhood density and frequency effects that have been studied in normals are extended to aphasic patients, and results are interpreted in terms of current models of language production.

According to the results obtained for normal subjects (Harley & Bown, 1998; Vitevitch, 1997; Vitevitch, ms in prep), neighbourhood variables are predicted to have a facilitative effect of speech production accuracy in aphasia. Thus, it is expected that low-frequency words will be more susceptible to error than high-frequency words, and that words from sparse neighbourhoods will be more susceptible than words from dense neighbourhoods. Facilitative effects of frequency and density would predict that words from low-frequency neighbourhoods will also be more susceptible than words from high-frequency neighbourhoods, although this result has not been found as consistently in normal error studies. If neighbourhood variables can be shown to influence the probability of a target being produced accurately, they might also be expected to have an effect on the outcome of errors. Outcome effects have not been strong in normal error studies, but it may be that the analyses were not sensitive enough to detect neighbourhood influences (see, for example Vitevitch, 1997). Furthermore, some outcome constraints, such as lexical frequency and syllable markedness, have been shown more consistently for aphasic than for normal errors (see Chapter 2). Thus, it is expected that, since low-frequency targets are less likely to be accurately produced, error outcomes will be higher in frequency than their targets (Blanken, 1990; Vitevitch, 1997). In addition, since targets from dense neighbourhoods are more likely to be accurately produced, error outcomes might be expected to come from more densely populated neighbourhoods.

Finding results for aphasic speakers which parallel results for normal speakers would support the continuity thesis, that is, the notion that aphasic deficits represent quantitative disruptions in normal processes, such as reduced efficiency of activation transmission, rather than qualitatively distinct processes (e.g. see Buckingham, 1999; Dell et al., 1997b). Corroborating findings would also provide further support for interactive theories of speech production (Vitevitch, ms in prep). On the other hand, if aphasic errors are found to be influenced by neighbourhood variables in ways that are different from normal errors, it would suggest the operation of pathological lexical access mechanisms, or perhaps of strategic compensatory processes not normally active in speech production.

Chapter 4. The Pilot Study

In conducting a study of aphasic errors, a number of "theory-laden" (Dell et al., 1997b) decisions must be made at each stage of the study, such as determining the context in which errors are collected, the classification of the errors, and the methods of analysis. As discussed in the previous chapter, there are many ways in which to accomplish these goals. In order to establish a motivated methodology for the main investigation, exploratory investigations were conducted on a different corpus of error data. This section describes these preliminary analyses, and the theoretical and methodological conclusions which helped to guide the main study.

Methodology

Subjects

Data for the pilot study was obtained from speech samples collected for a previous study (Gordon, 1998). In the original study, thirteen subjects were tested; all were native English speakers with a primary diagnosis of aphasia. No other criteria concerning type or severity of aphasia were used, so that the errors obtained could be considered representative of a range of types of aphasia, as seen in clinical settings. Of the thirteen subjects tested, three were excluded from the current analysis: one subject's aphasia was too mild (he made very few errors overall); one had a vocal tremor that made reliable transcription difficult; one had a degree of dysarthria which also made phonetic transcription unreliable. Characteristics of the remaining ten subjects are listed in Table 4-i (see following page).

Tasks

Subjects were tested on six of the expressive language tasks from the *Boston Diagnostic Aphasia Exam* (Goodglass & Kaplan, 1983), as listed in Table 4-ii

Table 4-i. Pilot Study: Subject Characteristics

Subject	Sex	Age[1]	TPO[2]	Fluency	Clinical Profile
P1	M	67	3	mixed	conduction
P2	M	52	1	fluent	Wernicke's
P3	F	80	6	fluent	conduction
P4	F	65	2	mixed	conduction
P5	F	63	5	mixed	transcortical motor
P6	F	79	8	mixed	anomic
P7	F	35	18	non-fluent	Broca's
P8	F	83	22	fluent	anomic
P9	F	77	54	fluent	anomic
P10	M	66	48	non-fluent	Broca's

[1] age in years at time of testing
[2] time post-onset in months

(following). These tasks varied in their stimulus presentation and response requirements, thus providing a range of contexts in which to collect errors. Tasks were presented in the same order (as listed) for all subjects. This is also the same order in which they occur in the BDAE, and follows a general continuum of decreasing structure, and thus of increasing difficulty for most aphasic patients.

Table 4-ii. Pilot Study Tasks

1) AS = Automatized Sequences (days of week & months of year only)
2) VA = Verbal Agility (speeded repetition of progressively longer words)
3) SR = Sentence Repetition (high & low frequency)
4) CN = Confrontation Naming (objects, actions & body parts only)
5) GN = Generative Naming (free naming of animals)
6) PD = Picture Description (Cookie Theft picture)

All tasks were audiotaped using a Sony Professional Walkman, and the entire samples were transcribed orthographically to provide the complete context in which errors occur. Lexical and phoneme level errors were transcribed using broad IPA transcription. Errors from all the tasks were included in the qualitative analyses, and in the statistical comparisons of errors to targets, but for the statistical comparisons of correct targets to error-targets, only errors from the Confrontation Naming and

Sentence Repetition tasks were used. These two tasks were chosen because they both represent a compromise in the trade-off between the degree of structure of the task, which makes targets easier to identify, and the naturalness of the task, which helps to ensure that errors are representative of speech errors occurring in spontaneous speech tasks. A list of the stimuli for these two tasks is provided in Appendix 4-i.

Qualitative Analyses

Lexical errors were extracted from the audiotaped speech samples, and were analyzed to assess the influence of their phonological neighbourhood characteristics. Errors were classified first according to their relationship (or relationships) to the intended targets: semantic, morphological, phonological, or unrelated. Semantic relationships included category members (e.g. *Friday* > *Monday*, *radio* > *television*), synonyms or near-synonyms (e.g. *captured* > *took*; *lid* > *hat*), associative relationships (e.g. *sink* > *tap*, *drinking* > *drive*), and 'functional substitutes', in which the substituted word has a meaning different from (but often associated to) the target, but fills the grammatical slot appropriately (e.g. *she* > *he*, *in* > *on*, *running* > *stopped*, *faucet* > *thing*, *smoked* > *ate*, *stool* > *stairs*). In short, any error which shared semantic features with its target was considered a semantic error. Morphological errors were those which differed from the target morphological form, but were otherwise semantically correct and phonologically accurate (e.g. *pry* > *pried*, *swallow* > *swallows*, *dripping* > *drippy*).

Because the focus of this study was on the phonological lexicon, semantic and morphological errors were excluded from further analyses. However, errors with both semantic and phonological relationships (i.e. mixed errors such as *swallow* > *sparrow*, *February* > *Friday*) were included because the phonological relationship is independent of the semantic relationship (whereas morphologically related items are by default also phonologically related). Furthermore, there is evidence that the existence of both types of relationship in one error is not merely coincidental (e.g. Dell & Reich, 1981), but is in fact due to the interaction of multiple connections within the lexicon. Some errors also bore a phonological relationship to a semantic substitution. Such errors differ from mixed errors in that they involve two separate stages of error (e.g. *cactus* > *picky*, with *prickly pear* as the intervening semantic substitution; *wrist* > /ɛŋksəl/, with *ankle* intervening). These were counted as two separate error processes, one semantic (e.g. *wrist* > *ankle*), and one phonological (e.g. *ankle* > /ɛŋksəl/).

Following several studies (e.g. Dell et al., 1997b; Gagnon et al., 1997; Roach, Schwartz, Martin, Grewal, & Brecher, 1996), error/target pairs were considered to be phonologically related if they shared at least one phoneme in the same syllable and word position (e.g. *curtains* > *coffer*, *possible* > *off*), or at least two phonemes in any syllable or word position (e.g. *Africa* > /næfə/, *phantom* > /mɑgən/). Unrelated errors were also included in the analysis, as long as the target could be determined. This allowed the inclusion of the numerous initial fragment errors which were subsequently self-corrected, providing validation of the intended target. Many of these occurred in sequences of phonemic approximation (e.g. /f-/, /fr-/, /k-/, /kon-/, /tʃail-/, /tʃainil/, *Chinese*). In addition, errors with clear phonological relationships to the context of the target, but not to the target itself, could also be included (e.g. *fled* > /sp-/, which is a perseveration of the previous word *spy*; *do* > *held*, which contains phonemes perseverated from *ahead*, a prior word in the sentence).

Phonological and unrelated errors were further classified according to the domain of the error, the type or mechanism of the error, its lexical status, and whether or not there was a contextual influence on the error. The domain of the error (i.e. whether it occurred at the word, syllable, cluster, phoneme, or feature level) did not factor into the statistical analyses, but did help to disambiguate the type of error that occurred. Types of errors counted included substitutions, additions, and deletions of words and phonemes. If a contextual source was evident, errors were further classified by the direction of the contextual influence—anticipation, perseveration, exchange, shift, or blend. Each error was counted separately, as long as it was deemed to be due to a separate mechanism. In this way, errors with contextual phonological sources could be separated from those with no apparent source. For example, in the sentence *The lawyer's closing argument convinced him*, the word *closing* was produced as /korzɪŋ/, an error which consists of the deletion of /l/ from *closing*, and the addition of /r/ perseverated from *lawyer's*. Subsequently, /korvɪŋ/ was produced, consisting of the perseveration of the syllable /kor-/, and the substitution of /v/ for /z/. The outcome of the error was classified according to its lexical status—word or non-word. Errors were considered real words if they could be found in a standard dictionary (*Webster's New World Dictionary*, Guralnik, 1986).

Because the domain over which context exerts an influence varies with the linguistic level at which the error occurs (Garrett, 1975), the definition of 'contextual' depended on the task and the level of the error. In the single-word production tasks (AS, VA, CN, GN), word-level perseverations included repetitions of any previous stimuli in the task; whereas phoneme-level perseverations were only counted if they came from within the current stimulus item or from the immediately preceding stimulus. In the sentence repetition task (SR), words anticipated or perseverated from the current sentence, or perseverated from the immediately preceding sentence, were counted as contextual. In the picture description task (PD) any word anticipated or perseverated from the description was counted as contextual. For both these tasks, phoneme-level anticipations and perseverations originated from anywhere within the word, or from the same syllable position in any other word within the sentence or utterance. The only feature-level errors which were counted were contextual accommodations of one phoneme to an immediately adjacent phoneme (e.g. *tip-top* > *tip-*/bɑp/).

This classification system allowed a qualitative analysis of the different patterns of results shown by different aphasics. Of particular interest were the relative proportions of semantic and phonological errors, and the types of contextual errors which predominated. Of particular interest for the statistical analysis was the breakdown of errors into contextual or non-contextual errors, and words or non-words, to assess whether these variables modified the effects of item frequency, neighbourhood density, and neighbourhood frequency.

Statistical Analyses

Phonological and unrelated errors were also analyzed statistically, to assess the effects of the three phonological neighbourhood variables, compiled in a lexicon of 20,000 words (see Luce & Pisoni, 1998). In this database, frequency of occurrence, obtained from Kučera and Francis (1967), was log-transformed. (For statistical purposes, the actual transformation involved replacing the zero frequencies with ones, log-transforming the data, then adding 1 to every transformed number, Vitevitch, 2000, personal communication.) Neighbourhood frequency, that is, the mean frequency of all of a word's neighbours, was similarly transformed. Neighbourhood density was defined as the absolute number of words sharing all but one phoneme with the target word.

There were two main types of comparisons conducted in the statistical analyses. First, targets of errors were compared to targets of correct utterances, in order to assess the effects of neighbourhood variables on the *susceptibility* of targets to being produced in error. Second, the errors themselves were compared to their respective targets, in order to assess the effects of neighbourhood on the *outcome* of the error. To make this distinction clear, some examples are listed in Table 4-iii (following). Of the four targets listed, two are produced in error (the

Table 4-iii. Examples of Error/Target Pairs

Errors	Targets
mauk/	mouse
✓	dog
/kæn/	cat
✓	horse

error is transcribed), and two are produced correctly (✓). For the first comparison, the targets of errors (*mouse* and *cat*) would be compared to the targets of correct utterances (*dog* and *horse*); for the second comparison, the errors (/mauk/ and /kæn/) would be compared to their targets (*mouse* and *cat*). Comparisons were carried out for each of the three variables by means of separate *t*-tests.

Results

Qualitative Analyses

Overall counts, for all subjects in all six tasks, yielded a total of 1130 error utterances. The total number of errors per subject, however, varied widely from only 8 errors to over 200 errors. Overall, the ratio of phonological errors to other types of errors (i.e. semantic or morphological) was 77.2% phonological to 22.8% non-phonological. Figure 4-i (page F-i) displays the relative proportions of these two error categories for each subject.

It was noted that subjects showed one of two predominant patterns: either they showed approximately equal numbers of phonological and non-phonological errors (apparent for subjects P2, P3, P5, and P8), or they showed many more (that is, more than 80%) phonological errors, compared to non-phonological errors. This pattern corresponded to some degree to the fluency diagnosis of the subject. Three of the four fluent aphasic subjects in the sample (P2, P3, and P8) showed the first pattern, with approximately equal numbers of phonological and non-phonological errors, whereas both of the non-fluent aphasic subjects in the sample (P7 and P10) showed the second pattern, with many more phonological than non-phonological errors. The other five subjects were diagnosed as mixed, and showed both patterns.

This finding suggests that non-fluent aphasic subjects exhibit a greater proportion of phonological errors; however, the subject sample in the present study is too small to draw any definitive conclusions. Furthermore, the diagnosis of fluency is a highly subjective method of classification, and must therefore be interpreted cautiously pending independent justification for the classification (Gordon, 1998). These results are consistent with previous findings which suggest that all sub-types of aphasic subjects show at least some phonological errors (Blumstein, 1973a; Goodglass et al., 1964; Mitchum et al., 1990).

Semantic and morphological errors were excluded from further analyses. All phonological and unrelated errors were classified as contextual (C) or non-contextual (NC), and as words (W) or non-words (NW). Overall, 51.6% of the errors were contextual, 48.4% were non-contextual; 51.5% were words and 48.5% non-words. Individual subjects showed a range of proportions for both classifications, neither of which bore any apparent relation to the subjects' fluency diagnoses. However, all subjects showed at least 30% NC errors, a significant proportion which serves as some justification for granting them more attention in speech error research.

Contextual errors (n=499) were further classified according to the direction of contextual influence: anticipatory, perseveratory, shift, exchange, or blend. The breakdown of these types is illustrated in Figure 4-ii (page F-i). This analysis was motivated by findings that, for normal subjects, anticipations have been shown to be more frequent than perseverations (e.g. Garnham et al., 1981), whereas aphasic subjects tend to show more perseverations than anticipations (e.g. Schwartz et al., 1994). The ratio of anticipations to perseverations (hereafter A:P ratio) was also one of the variables that Dell and colleagues (1997b) found could be modeled by making quantitative alterations in the parameters of a computerized lexical network (i.e. by weakening connection strength), thus illustrating the continuity between aphasic and normal patterns of errors (the 'continuity thesis').

In this study, the majority (70%) of errors were perseveratory, followed by 22% anticipatory errors, which corresponds to the pattern observed by Dell and colleagues (1997b). These investigators also found that when error rate is higher, errors tend to be more perseveratory; thus, they hypothesized that the A:P ratio reflects the severity of the deficit giving rise to the errors. A severity relationship was also suggested here, by looking at the proportions of contextual errors shown by individual subjects (see Figures 4-iii and 4-iv).

As evident in Figure 4-iii (p. F-ii), most subjects showed the same pattern found by Dell et al., with high proportions of perseverations relative to anticipations. Some subjects, however, (P3, P5, P8, and P9), did not (see Figure 4-iv, p. F-iii). Of these four subjects, three were fluent aphasics, two of whom were diagnosed as anomic, which is often the mildest type of aphasia. More specifically, these four subjects also had the four lowest error rates of the ten subjects, ranging from 35 down to only 2 contextual errors. Thus, these results provide support for the relationship of A:P ratio to severity, and for the continuity thesis (Dell et al., 1997b; Martin et al., 1994; Schwartz et al., 1994).

Statistical Analyses

For all the statistical analyses, neighbourhood values were obtained from the on-line lexicon described earlier. Some items, however, were not available in the lexicon, and had to be excluded from the analyses. The proportion of exclusions was monitored in each analysis to ensure that there were not significantly more items excluded from any one cell of an analysis, which could potentially bias the results.

Target Susceptibility

In the first analysis, targets produced in error were compared to targets produced correctly, in order to assess the influence of phonological neighbourhood factors on the susceptibility of items to error. As a preliminary step, neighbourhood values for targets of contextual and non-contextual errors were compared to see whether they differed. If so, this would suggest that they should be analyzed separately.

Contextual *vs* Non-Contextual Errors: In this analysis, 16.3% of contextual items and 20.6% of non-contextual items had to be excluded because they were not found in the lexicon. Results of the *t*-test comparisons are presented below in Table 4-iv. Contextual error targets were not significantly different from non-contextual error targets on any of the three variables—word frequency (log frequency), neighbourhood density (N density), or neighbourhood frequency (log N frequency). Therefore, contextual and non-contextual error targets were pooled in subsequent analyses of error targets to correct targets.

Table 4-iv. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Contextual (C) *vs* Non-Contextual (NC) Errors

Dependent Variable	C Errors (n=438)	NC Errors (n=390)	*t*-test (p<0.05)
Log Frequency	2.221	2.338	n.s.
N Density	9.406	10.326	n.s.
Log N Frequency	1.382	1.428	n.s.

Target Susceptibility (SR): All targets which were correctly produced were compared to all targets produced in error, for the two tasks of sentence repetition (SR) and confrontation naming (CN). The ratio of error targets to correctly produced targets yielded an error rate of 32.6%. For each individual error, the target was counted once; thus, a target which contained two separate error processes was counted twice. In the SR task, 12.3% of correct targets and 10.1% of error targets were not found in the lexicon. As illustrated in Table 4-v, correct-targets were found to be significantly different from error-targets on all three variables. That is, correct targets were higher in frequency of occurrence, had a greater number of phonological neighbours, and came from higher frequency neighbourhoods than targets of errors. These results suggest that a larger, higher frequency phonological neighbourhood actually *facilitates*

Table 4-v. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Correct (C) Targets *vs* Error (E) Targets (Sentence Repetition)

Dependent Variable	C Targets (n=648)	E Targets (n=313)	*t*-test (p<0.05)
Log Frequency	4.124	2.530	*
N Density	13.995	11.061	*
Log N Frequency	2.331	1.598	*

production, supporting previous studies of normal error production (Harley & Bown, 1998; Vitevitch, 1997; Vitevitch, ms in prep), a finding that has been referred to as the 'conspiracy effect' (Taraban & McClelland, 1987; Vitevitch, 1997).

Target Susceptibility (CN): The ratio of error targets to correct targets yielded an error rate of 56%. All stimuli in the confrontation naming task were found in the lexicon. Results of the *t*-tests comparing mean neighbourhood values are shown in Table 4-vi.

Table 4-vi. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Correct (C) Targets *vs* Error (E) Targets (Confrontation Naming)

Dependent Variable	C Targets (n=171)	E targets (n=218)	*t*-test ($p<0.05$)
Log Frequency	2.299	2.083	*
N Density	11.918	10.249	n.s.
Log N Frequency	1.754	1.670	n.s.

As in the SR task, word frequency was significantly higher for correct targets than for error targets. However, unlike the SR task, no significant differences were found between correct and error targets on either neighbourhood density or neighbourhood frequency, although the differences were in the same direction as in the SR task. These null effects may be explained by differences between the two tasks. One of the most obvious differences is that the stimuli in the SR task include both content and function words, whereas the CN task contains only content words (nouns and verbs). To investigate whether this was the factor that gave rise to the discrepant findings, target words from the SR task were divided into content and function words, and separate *t*-tests were run for each set of stimuli.

Target Susceptibility (SR—Content *vs* Function Words): The error rate was 48.3% for content words, but only 5.6% for function words, which clearly indicates a difference in error susceptibility between the two classes of words, as has been noted previously (e.g. Buckingham & Kertesz, 1974; Butterworth, 1979; Garrett, 1980). Nevertheless, the significant effects of the first neighbourhood analysis appear to be maintained for both content and function words. The results of these analyses are displayed in Tables 4-vii and 4-viii. For the function words, however, the effect of

Table 4-vii. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Correct (C) Targets *vs* Error (E) Targets (Content Words)

Dependent Variable	C Targets (n=314)	E Targets (n=293)	*t*-test ($p<0.05$)
Log Frequency	2.904	2.400	*
N Density	15.503	10.648	*
Log N Frequency	1.882	1.546	*

Table 4-viii. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Correct (C) Targets *vs* Error (E) Targets (Function Words)

Dependent Variable	C Targets (n=334)	E Targets (n=20)	*t*-test ($p<0.05$)
Log Frequency	5.271	4.427	*
N Density	12.578	17.100	*
Log N Frequency	2.754	2.354	*

neighbourhood density is reversed; that is, neighbourhoods are more dense for errors than for targets, suggesting an inhibitory effect of neighbourhood density. (It should be noted, though, that the results of the function word analysis are based on only 20 error targets in comparison with over 300 correct targets; therefore, all of these results should be treated with caution.) On the other hand, the content word analysis, which is the more appropriate comparison to the CN task, still shows clear facilitative effects of frequency, neighbourhood density and neighbourhood frequency.

The question remains, then: If the discrepant results are not due to the grammatical class of the stimuli, what might be the reason behind the discrepancy?

One possibility is a difference in the number of items in the two tasks. In the CN task, the analysis is based on 389 utterances, counting both correct and error targets, but these utterances were in response to only 18 different stimulus items. In the SR task, on the other hand, the analysis is based on 961 utterances, in response to 98 different stimulus items, counting each target word in each sentence. With so few different items in the CN task, it may be that the range of densities of the targets (from 0 to 31 neighbours) was not sufficient to show an effect. Perhaps a more extensive naming task, then, would show neighbourhood differences.

Alternatively, the different stimulus and response characteristics of the two tasks might influence the susceptibility of the targets to error. In the SR task, the phonological stimulus is provided by the examiner, and the subject is required to repeat it, whereas in the CN task, the subject is required to retrieve the name of the picture presented, and to encode it phonologically. Because the two tasks involve different cognitive processes, it is possible that the differences may be attributable to processes specific to the tasks. For example, the repetition task involves a receptive component which the naming task does not; perhaps neighbourhood effects are exerted at this stage. However, this seems unlikely given that, if neighbourhood effects arise at the auditory perceptual stage, subjects would be expected to show signs of difficulty in recognizing the stimuli, and errors would show evidence of perceptual confusions, neither of which were observed. Nevertheless, the possibility remains open that task-related artifacts influenced the results.

Error Outcome

Given the facilitative effects of neighbourhood density and frequency that seemed to affect the susceptibility of stimulus items, at least in the SR task, it was expected that the errors that were actually produced would also have higher density and higher neighbourhood frequency values than the targets that were supposed to be produced. In the error/target comparisons, only whole-word errors were included because the lexical status of fragments cannot be determined, but errors from all the tasks were pooled together.

Word Errors *vs* Non-Word Errors: As a preliminary step, word and non-word errors were compared to see whether their neighbourhood characteristics differed. Note that the word/non-word distinction is between errors; all targets are, of course, real words. It should also be noted that frequency of occurrence is not a relevant variable for non-words because they do not exist in the lexicon. Thus, no frequency analyses were done whenever the data in the comparison included non-words. Results of the word/non-word comparison, presented in Table 4-ix, illustrate that both neighbourhood characteristics differ significantly between word and non-word errors. Word errors are significantly higher in neighbourhood density and neighbourhood frequency than non-word errors, suggesting that the two sets of data should be analyzed separately. Another reason for analyzing them separately is to allow assessment of a frequency effect for the word errors.

Table 4-ix. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Word (W) *vs* Non-Word (NW) Errors

Dependent Variable	W errors (n=240)	NW errors (n=243)	*t*-test ($p<0.05$)
Log Frequency	2.481	(n/a)	(n/a)
N Density	13.410	4.297	*
Log N Frequency	2.234	1.260	*

Error Outcome (W Errors): In the comparison of word errors to their targets (see Table 4-x), errors were not found to have a significantly higher mean frequency than their targets. This lack of an effect was somewhat surprising, given the robustness of other frequency effects in the lexical access literature (e.g. Dell, 1990; Favreau et al., 1990; Stemberger & MacWhinney, 1986; Williams & Canter, 1982), but it may reflect the fact that many of the word errors here occurred by chance. Word errors may be produced either by the substitution of one word for another, which are true word errors, or by the substitution of one or more phonemes in the target which, by chance, creates a different real word. In order to discriminate between these possibilities, it would be necessary to estimate what the chance rate of word production is, and to determine whether it is exceeded by the obtained incidence of word error production. However, although this would indicate whether or not true word substitution errors exist in the corpus, it would not identify which errors constitute word substitutions, and which are words by chance, so the two types still could not be disambiguated. The effect of neighbourhood density was also not significant for word errors, but a significant neighbourhood frequency effect was found, where errors came from higher frequency neighbourhoods than their targets, as expected.

Table 4-x. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Word (W) Errors *vs* Word-Error (WE) Targets

Dependent Variable	W Errors (n=240)	WE Targets (n=240)	*t*-test (p<0.05)
Log Frequency	2.481	2.489	n.s.
N Density	13.410	13.350	n.s.
Log N Frequency	2.234	1.776	*

Error Outcome (NW Errors): For non-word errors (see Table 4-xi), unlike word errors, no significant effect of neighbourhood frequency was found, but a significant neighbourhood density effect was found. The latter, however, was in the unexpected direction; that is, errors came from lower density neighbourhoods than their targets. This result may be explained by the extent of deviance of some of the non-word (neologistic) errors. Recall that these analyses included errors which were unrelated to their targets, and that phonological relatedness was also defined quite liberally. Thus, many of the non-word errors were extremely un-word-like and would be expected to have few, if any, neighbours. For example, one subject produced the neologisms /lifənblɪpər/ for *dining room*, and /pɑtərwɪlsə/ for *caterpillar*.

Table 4-xi. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Non-Word (NW) Errors *vs* Non-Word Error (NWE) Targets

Dependent Variable	NW Errors (n=243)	NWE Targets (n=243)	*t*-test (p<0.05)
Log Frequency	(n/a)	2.141	(n/a)
N Density	4.297	8.015	*
Log N Frequency	1.260	1.148	n.s.

Error Outcome (C *vs* NC Errors): Although no consistent neighbourhood effects were found to influence the outcome of errors to this point, an effect was expected in the comparison of contextual to non-contextual errors. Because the outcome of contextual errors can, by definition, be explained by the phonological context, it was expected that non-contextual errors, for which no such explanation can be found, might exhibit a greater influence of phonological neighbourhood characteristics. For this reason, errors and targets were compared separately for contextual and non-contextual errors. Results, presented in Tables 4-xii and 4-xiii,

Table 4-xii. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Contextual (C) Errors *vs* Contextual Error (CE) Targets (Contextual Errors)

Dependent Variable	C Errors (n=262)	CE Targets (n=262)	*t*-test (p<0.05)
Log Frequency	(n/a)	2.247	(n/a)
N Density	8.693	10.087	n.s.
Log N Frequency	1.662	1.413	*

Table 4-xiii. Mean Values for Frequency, Neighbourhood Density, and Neighbourhood Frequency: Non-Contextual (NC) Errors *vs* Non-Contextual Error (NCE) Targets

Dependent Variable	NC Errors (n=216)	NCE Targets (n=216)	*t*-test (p<0.05)
Log Frequency	(n/a)	2.404	(n/a)
N Density	8.198	11.424	*
Log N Frequency	1.756	1.513	n.s.

showed that, indeed, the effect of density was significant for non-contextual errors only. However, the effect was again opposite to that predicted, such that neighbourhood density was higher for targets than errors.

Summary and Discussion

To summarize the results: qualitative analyses suggested that the proportion of phonological to non-phonological errors may be related to the fluency of speech production in aphasic subjects, although patterns of word and non-word errors, and contextual and non-contextual errors, showed no clear relationship to clinical sub-types of aphasia. Patterns of contextual errors, specifically the ratio of anticipations to perseverations, were found to correspond to aphasic patterns shown in the literature, and seem to be related to severity in aphasia. Because of the small sample size, and the lack of any statistical comparisons, these conclusions must be considered tentative.

In the statistical analyses, phonological neighbourhood variables appeared to exert a facilitative effect on the susceptibility of target items to error, consistent with the conspiracy effects found in error studies with normal subjects (Vitevitch, 1997; Vitevitch, ms in prep). That is, those targets which were produced correctly were significantly higher in frequency, and had higher neighbourhood densities and higher neighbourhood frequencies than those targets produced in error. This result was clearly shown in the sentence repetition task, at least for content words, but was not found in the confrontation naming task, possibly because there were too few stimulus items, without a sufficient range of density values, or because of task-specific differences.

Despite the neighbourhood effects on target susceptibility (i.e. comparing correct targets to error targets), no consistent neighbourhood effects were found on error outcome (i.e. comparing errors to targets). Word errors showed no effect of word frequency or neighbourhood density relative to their targets, although an effect of neighbourhood frequency was shown. No effect of neighbourhood frequency was found for non-word errors, and the neighbourhood density effect was the reverse of what was expected—non-word errors came from less dense neighbourhoods than their targets. The comparison of effects for contextual and non-contextual errors was similarly inconsistent. The effect of neighbourhood density was significant for non-contextual errors only, but in the opposite direction from predicted—targets came from denser neighbourhoods than errors. An effect of neighbourhood frequency in the expected direction was found, but only for contextual errors, not non-contextual errors. Thus, while neighbourhood factors seem to influence the likelihood that an item is produced in error, they do not seem to have a consistent effect on the outcome of the error.

Implications for the Main Study

The main finding of this preliminary analysis is that aphasic errors appear to show the same pattern of neighbourhood effects in production tasks as do normal errors; that is, that a denser, higher frequency neighbourhood facilitates correct production. Although the facilitative effect of the phonological neighbourhood is suggestive, inconsistencies in these results remained to be addressed. It was decided that, for the main study, a more extensive confrontation naming task would be used in hopes of illustrating neighbourhood effects that were not evident here. In addition, a spontaneous speech task would be conducted to provide a corpus of errors which would be more representative of natural speech. The lack of significant differences between contextual and non-contextual errors suggested that this is not an important distinction to make in analyzing the effects of neighbourhood variables on error production; therefore, this distinction was not made in the main study.

Other methodological considerations concern the way in which errors and targets were counted. As noted in the description of the methodology, every phoneme error was counted separately, and therefore the targets they occurred in were counted more than once. The size of the domain of phonological errors relative to other types of errors means that this method has the effect of inflating the proportions of phonological errors relative to other types of error. That is, several phonological errors may occur in the same word, but only one semantic substitution may occur at a time for a given word. Furthermore, this method makes the implicit assumption that multiple phonological errors within a word occur independently, which is probably rarely the case. Therefore, in the main study, each error utterance was counted as a single error. Because contextual effects were not to be examined in the main study, the need to differentiate contextual from non-contextual phonological errors (the reason behind the method used in the pilot study) does not arise. Another potential problem in the way in which errors were defined involves the inclusion of unrelated errors. Although cues such as self-corrections were used to identify the targets of such errors, their targets can never be determined with the same confidence as for related errors. Therefore, the main study included only phonologically related errors.

Another potential methodological short-coming of the pilot study is the method of identifying correct targets and error-targets. Here, correct targets in each task included every stimulus item which was produced correctly by a subject, whereas error-targets included every stimulus item which was produced incorrectly. Given the limited range of stimuli in the two tasks, the two sets of targets consist of essentially the same items, but in different relative numbers. For example, *key* and *hammock* were both produced correctly by some subjects and incorrectly by others, so *key* might have appeared three times in the error-target corpus and seven times in the correct-target corpus, whereas *hammock* might have appeared nine times in the error-target corpus and once in the correct-target corpus. This overlap may have diminished the likelihood of finding an effect, especially in the CN task, which had so few stimuli. As will be seen in the description of the main study, it was determined that a more valid method would involve comparing the error-targets, in their relative numbers, to an independent set of control targets which were never produced in error.

Finally, the pilot study used data collected for a different experiment. However, because the data in the main study was collected specifically for the purposes of this study, it provided a more controlled sample of errors to test the effects of neighbourhood variables on aphasic error production.

Chapter 5. The Main Study

The principal investigation was designed to investigate the role of phonological neighbourhood variables on error production in aphasia. Based on findings from the pilot study, errors were elicited using a confrontation naming task and a spontaneous speech (picture description) task. These tasks were chosen primarily because they require no receptive language processing, and they provide the subject with no clues about the target's phonological structure. Aphasic subjects were unselected by clinical sub-type or severity, as in the pilot study, but more subjects were tested to add power to the experiment. Where possible, more powerful statistical methods were used in the main study than were used in the pilot study. For example, the effect of the lexical status of the error was incorporated into analyses of variance, rather than doing separate *t*-tests. In addition, the accuracy of naming was used as the dependent variable in a multiple regression analysis. It was hypothesized that frequency, neighbourhood density and neighbourhood frequency will influence the susceptibility of targets to error in a facilitative manner, as has been found for normal subjects (Harley & Bown, 1998; Vitevitch, 1997; Vitevitch, ms in prep), and in line with the pilot data. Thus, targets with higher frequency, neighbourhood density, and neighbourhood frequency values were expected to be more likely to be produced correctly. In addition, errors were hypothesized to be more frequent than their targets, and to come from higher-density, more frequent neighbourhoods, although null results in this analysis would not be surprising, given results of previous studies and the pilot study.

Methodology

The Subjects

Aphasic subjects were recruited from several rehabilitation hospitals in the Montréal, Québec area and an out-patient day-centre in the York-Durham region of Ontario. Forty-three native-English-speaking subjects with a primary diagnosis of aphasia were tested, three of whom had previously participated in the pilot study. All were at least three months post-onset of their stroke at the time of testing. A battery of language tests (the auditory comprehension and oral expression tasks from the *Boston Diagnostic Aphasia Exam*, (BDAE, Goodglass & Kaplan, 1983) was administered to each subject prior to the experimental tasks to ensure that visual-perceptual abilities, hearing acuity and auditory comprehension were adequate to perform the tasks, and to provide general information on the subjects' aphasic deficits. On the basis of their performance on these tasks, seven subjects were excluded from the experimental tasks due to the severity of their aphasic or dysarthric deficits. Among the 36 subjects left, 15 were male and 21 were female. Their mean level of education ranged from 6 to 16 years, with an average of 11.4 years. The average age of the subjects at the time of testing was 68 years, and ranged from 47 years to 86 years. The mean post-onset time for the group was 51 months at the time of testing, ranging widely from 4 to 132 months. A complete list of the aphasic subjects tested is presented in Table 5-i (see following page).

A small group of six non-brain-damaged elderly subjects was also tested on the picture description task to pre-test the stimuli and to provide a corpus of normal speech. These individuals were recruited from a pool of adult volunteers, and were paid a small amount for their participation. Control subjects were native English speakers, with no

Table 5-i. Aphasic Subjects Tested in the Main Study

Subject	Sex	Educ. (years)	Age @ Testing (years) NR[1]	Age @ Testing (years) PNT[2]	Time Post-Onset (months) NR[1]	Time Post-Onset (months) PNT[2]
M01	m	7	70	71	28	42
M02	f	11	N/A	67	N/A	41
M03	f	11	81	N/A	33	N/A
M04-M07	excluded					
M08	f	16	49	50	29	45
M09	m	14	67	68	16	28
M10	f	12	81	82	39	57
M11	f	12	55	N/A	59	N/A
M12	m	14	50	51	121	132
M13	f	9	75	76	19	28
M14	f	11	84	85	77	84
M15	f	9	86	86	19	24
M16	m	9	65	65	8	12
M17	m	14	68	68	6	10
M18	f	13	75	75	4	12
M19	m	9	79	79	38	41
M20	f	8	82	82	58	61
M21	f	9	67	67	62	63
M22	f	12	71	71	63	64
M23	excluded					
M24	m	11	80	80	8	6
M25	f	9	64	64	14	14
M26	f	14	N/A	47	N/A	80
YD01	excluded					
YD02	m	16	49	49	66	66
YD03	excluded					
YD04	m	unknown	60	60	46	46
YD05	m	6	53	53	129	129
YD06	f	12	85	85	>3	>3
YD07	m	10	77	77	86	86
YD08	f	unknown	52	52	49	49
YD09	f	12	72	72	26	26
YD10	m	16	52	52	11	11
YD11	m	16	69	69	109	109
YD12	f	unknown	58	58	47	47
YD13	m	11	61	N/A	31	N/A
YD14	f	8	75	75	15	15
YD15	m	13	68	68	107	107
YD16	f	12	68	68	119	119
YD17	f	10	81	81	45	45
Mean		**11.4**	**68.5**	**68.3**	**48.1**	**53.1**
Max		**16.0**	**86.0**	**86.0**	**129.0**	**132.0**
Min		**6.0**	**49.0**	**47.0**	**4.0**	**6.0**

[1] NR = Norman Rockwell picture description task
[2] PNT = *Philadelphia Naming Test*

significant hearing impairment, and no history of stroke or other brain injury. There were three men and three women in the control group, with a mean level of education of 14 years (range: 11 to 18 years). At the time of testing, they ranged from 62 to 86 years old, with a mean of 72 years. A list of the control subjects follows in Table 5-ii.

Table 5-ii. Control Subjects in the Main Study

Subject	Sex	Education (years)	Age @ Testing (years)
C1	m	11	62
C2	f	18	64
C3	f	12	72
C4	m	12	77
C5	m	16	75
C6	f	15	86
Mean		**14.0**	**72.7**

The aphasia profiles obtained from the BDAE were subsequently used to provide the necessary background information for interpreting individual error patterns, but were not used to group subjects. Nor were there any other exclusionary criteria in terms of type or severity of aphasia set *a priori*, so that the errors obtained might be considered a representative sample of the range of types and frequencies of errors produced in natural communicative situations by an unselected group of aphasic subjects. Other studies have excluded non-fluent subjects because of the possibility that their phonological errors might be attributable to articulatory execution stages of production, rather than phonological planning (e.g. Dell et al., 1997b; Martin et al., 1996). However, studies have also shown that fluent and non-fluent subjects cannot be clearly distinguished by their patterns of phonological errors (e.g. Blumstein, 1973a; Goodglass et al., 1964). (This is *not* to claim that fluent and non-fluent subjects do not show differences in phonological processing—numerous studies attest to this fact—but simply that it is, as yet, unclear that such a gross distinction as 'fluent/non-fluent' corresponds in any systematic way to a distinction between errors of phonological encoding and errors of phonetic implementation (Blumstein, 1991; Gordon, 1998).) Nickels and Howard (1995) note that excluding apraxic subjects "makes a number of assumptions that may not be justified, not least that the deficits [of apraxics and fluent aphasics] are indeed separable and distinct rather than points on a continuum" (p. 220). Both types were accepted for the present study, to avoid biasing results with *a priori* assumptions.

The Tasks

Because we are unable to directly observe the linguistic processes under investigation, it is important to rely on data collected in a variety of experimental contexts, which differ, for example, in the naturalness of the task, the constraints of the vocabulary elicited, and the characteristics of the stimuli. Investigations into the mechanisms of normal language production have relied on both naturally occurring speech errors and experimentally elicited errors produced by non-brain-damaged and aphasic subjects (Blumstein, 1973a; Bock & Levelt, 1994; Dressler, 1979; Garrett, 1980). Spontaneously produced errors are of interest here because they represent processes that occur in natural speech production. However, it was anticipated that large samples would be required to collect a sufficient number of errors, that many of the errors would be difficult to transcribe accurately, and that targets might be difficult to determine in an unstructured task. Therefore, a more structured task of picture naming was included to supplement the spontaneous speech error corpus, and to provide a more controlled set of stimuli.

Norman Rockwell Picture Description Task (NR)

Stimuli: In the only other group study of aphasic speech errors based on a corpus of spontaneous speech, Blumstein (1973b) used an interview format. Here, however, a picture description task was chosen to somewhat delimit the vocabulary used by subjects, so that the intended targets would be easier to determine, and to minimize the potential influence of experimenter input on the subjects' output. Subjects were asked to describe the scene shown in a number of Norman Rockwell (hereafter, NR) prints chosen for their ability to stimulate discussion. To help select the stimulus pictures, 25 NR prints were presented to the six control subjects. These prints were chosen according to general criteria of visual clarity, emotional content and topic relevance to the age-group tested. For example, several of the pictures dealt with war-time themes, which were expected to have particular personal relevance for this age-group of subjects. Mean word counts were calculated across the six subjects for each picture, as a gross measure of speech output, and ranged from 747 words to 2653 words per picture. Of the 15 pictures which inspired the most output, ten were chosen for presentation to the aphasic subjects, according to their variety of subject matter and humorous content. Black and white reproductions of the ten selected pictures are shown in Appendix 5-i.

Subjects: Of the 36 subjects who passed the BDAE pre-testing, two did not complete the picture description task: one chose to discontinue the task because of her severe non-fluency, and one was unavailable to complete the testing. Two of the samples collected were difficult to transcribe accurately due to background noise, but these two subjects were re-tested. With these exclusions, 34 samples of the NR picture description task remained.

Procedure: Examiners administered the tests in an environment that was as quiet and free of auditory and visual distractions as possible, although to some extent, this was beyond the examiners' control, as most of the subjects were tested in their own homes. However, any environmental influences were noted, such as a family member entering the room and speaking to the subject, or the telephone ringing. When such interruptions occurred, the task was halted until the distraction was gone. Examiners made notes during the administration of the task which were used during transcription to help disambiguate targets, for example, by noting what part of the picture the subject was pointing to when producing an error.

Examiners were advised not to intervene during the subjects' responses, and not to provide any cueing. However, non-specific prompting, such as "Is there anything else?" or "What else do you see?" was allowed in order to encourage more output. These prompts were intended to minimize any strategic differences between subjects in the way the task was performed. For example, in the face of word-finding difficulties, some subjects might abort the attempt (e.g. "I don't know"; "There's nothing happening in this picture") rather than risk making an error. Subjects who tended to simply list objects in the picture were guided with prompts such as "What is happening?" or "What is the story in this picture?" Despite these precautions, cueing and feedback were sometimes provided in order to preserve rapport with the subject or to maintain the subject's attention. In all such cases, errors that were made following cueing or feedback were excluded from the analyses.

The ten pictures were presented in the same order for all subjects, and all pictures were administered in the same session.

Philadelphia Naming Task (PNT)

Stimuli: To expand and replicate the corpus of spontaneously occurring errors, and to provide greater control over the targets, the *Philadelphia Naming Test* (PNT, Roach et al., 1996) was administered to the aphasic subjects. This test includes 175 line drawings of objects, with names varying in length from one to four syllables and varying in frequency of occurrence from 1 to 2110 per million, based on Francis and

Kučera's (1982) noun frequencies. A list of the PNT stimulus items is provided in Appendix 5-ii.

Subjects: Of the 36 subjects remaining in the experimental pool, two were unavailable to complete the PNT, and the examiner discontinued the task for another subject because he began to perseverate, using the same stereotyped utterance on every trial. As in the NR task, an additional two of the samples were unintelligible due to background noise, but these subjects were retested. A further subject was excluded because he intentionally mispronounced many of the stimuli for comic effect. In all, 32 PNT samples remained for analysis.

Procedure: Although the PNT can be administered on a computer in order to measure reaction times of naming responses, reaction times were not relevant to this study, so the pictures were presented on paper, one by one. Examiners were instructed to allow subjects plenty of time for their initial response, as well as time to repair or revise their response. It was assumed that if the task were 'self-timed' rather than imposing time constraints, subjects' responses would reflect more natural word-retrieval processes. If subjects showed a tendency to give up quickly, they were encouraged to guess, but examiners were asked not to provide cues or feedback regarding the accuracy of responses. Non-specific encouragement (e.g. "Ok", "Good try", "That's it") was allowed, regardless of the accuracy of the response. As in the NR task, cueing and feedback were sometimes provided in order to maintain the subjects' cooperation, or to try to prevent perseveration of an item (Gagnon et al., 1997), but any error responses given following such cues or feedback were not counted.

PNT pictures were presented in the random order dictated by the test protocol to all subjects. Like the NR task, the PNT was completed in one session, although some subjects did the two tasks in separate sessions.

Transcription Procedure

Both tasks were tape-recorded using a Sony Professional Walkman WM D6-C, and speech samples were transcribed using a Sony BM-75 dictator/transcriber. The entire speech sample elicited for each task, even commentary which did not pertain directly to the picture being named or described, was transcribed orthographically in order to provide the full context in which the errors occurred (Fay & Cutler, 1977; Stemberger, 1985). This not only enabled the experimenter to trace influences of the phonological context surrounding errors, but also to keep track of (and exclude) any errors which were influenced by environmental intrusions, or by cues or feedback from the examiner. Phonological errors, both words and non-words, were transcribed using the broad phonetic (i.e. phonemic) transcription system of the International Phonetic Association. Appendix 5-iii lists the IPA symbols used and their descriptions.

The reliability of the transcription was ensured through a rigorous process of consensus among three transcribers, all of whom were experienced in the use of phonetic transcription and were familiar with the characteristics of aphasic speech. Following some previous studies (e.g. Christman, 1994; Gagnon et al., 1997; Kohn et al., 1998), the first two (T1 and T2) transcribed independently, and the third (T3) resolved discrepancies. T2 and T3 were naive to the purpose of the experiment (T1 was the author). The speech samples were transcribed in their entirety by one of the examiners, usually the one who conducted the testing for that subject, but the phonetically transcribed portions of all of the samples were checked for consistency by one examiner (T2). Independently (i.e. by listening to the audiotape without reference to T2's transcription), the author (T1) transcribed all of the word-level and phoneme-level errors. The two independent transcriptions (T1 and T2) of each error were compiled and compared by the author, and any errors missed by one transcriber were recorded at this stage. Next, a third listener (T3) compared the transcriptions of each error, and adjudicated between the remaining discrepancies. At each stage, the transcriber listened to the original tape-recording, re-playing errors as many times as necessary to make a reliable transcription. Thus, each transcriber had access to the context of utterance, which helped to make each transcription as accurate as possible.

At the final stage, T3 was asked to make a 'reasonably confident' decision, for each discrepancy, among the following three options: a) T1 is accurate; b) T2 is accurate; or c) neither T1 nor T2 is accurate, and T3 records a different transcription. Failing this, T3 chooses a fourth option: d) no decision can be made with reasonable confidence. This fourth option was provided as a conservative measure, so that truly ambiguous utterances would not have to be included through a forced-choice procedure. The results of the reliability assessment are presented in the next section.

Some discrepancies were considered irrelevant, such as differences in the use of unstressed vowels (e.g. /pʌmpkɪn/ *vs* /pʌmpkən/); in the transcription of a flap (e.g. /bʌtər/ *vs* /bʌdər/) or in the transcription of affrication in certain environments (e.g. /tri/ *vs* /tʃri/). The use of symbols was determined to some extent by the phonetic transcriptions used in the neighbourhood lexicon. For example, because there is no flap in the phonetic symbol system of the neighbourhood database, the orthographically appropriate stop (i.e. /t/ or /d/) was used instead. The phonetic symbols used in the neighbourhood database are listed beside the corresponding IPA symbols in Appendix 5-iii.

A consensus was required on not only the identity of the phonemes in each item, but also on whether off-target items resulted from phonetic distortion, normal co-articulatory processes, dialectical variation from standard pronunciation, or accent effects, as opposed to clear phonemic errors. In addition, transcribers had to agree on the identity of the target, and on whether each item was complete or incomplete. Incomplete items, or fragments, were subjectively judged on the basis of auditory cues such as segment duration, intonation and pausing, as described in the *Philadelphia Naming Test* (Roach et al., 1996). To be included, fragments were defined as consisting of at least one consonant and one non-schwa vowel (Roach et al., 1996). In order to make decisions with a reasonable level of confidence, T3 was instructed to be conservative: utterances that could easily be perceived in more than one way were to be judged ambiguous; judgements involving whether the utterance was correct or incorrect should be biased towards 'correct'; similarly, judgements involving whether the error involved a phonemic substitution or a phonetic distortion should be biased towards the 'distortion' interpretation. That is, only utterances confidently perceived by two of three listeners to be unambiguous phonemic errors were counted.

The Analyses

Error Classification

The set of errors defined through the reliability procedure was further pruned by the elimination of errors which were determined by consensus not to be phonologically related errors after all (e.g. correct productions, distortions, dialectical variations). At this stage, immediate repetitions of the same error, and fragment errors which were repeated in subsequent expansions were also eliminated. For example, if a subject produced /kɑn- kænəl- kænəldər/ for *calendar*, the first fragment was counted, as it was subsequently revised, but not the second, as it was subsequently repeated within the expansion; if /kæn- kænəl- kænəldər/ were produced, only the final attempt was counted. Similarly, repeated perseverations and stereotypical utterances were counted only the first time they appeared. The final error set included only phonologically related errors. Although some of the error/target pairs were related in other ways (e.g. mixed errors, contextual word substitutions, perseverations), these were retained only if they were also phonologically related.

Phonological Relatedness: As in the PNT (Roach et al., 1996), errors were considered to be phonologically related to their targets if they matched minimally on one phoneme occurring in the same syllable and word position, or two phonemes occurring in any position. Although this may seem to be a very liberal definition of phonological relatedness, it is theoretically motivated by Dell's (1986) interactive activation model of lexical access, in which activation spreads among phonologically related words through their shared phonemes. Only one phoneme overlap is required for activation to spread from one word to another, although a greater degree of overlap would increase the activation of a phonological neighbour.

PNT Coding: Responses on the *Philadelphia Naming Test* were coded according to the protocol described by Roach et al. (1996), even though most of the error categories were ignored for the present study. A sample of the score sheet is provided in Appendix 5-iv. Up to three responses on each item are coded: the initial response, consisting of either a fragment or complete utterance; the first complete attempt (if the initial attempt is a fragment); and the final complete attempt. Furthermore, responses are coded at two levels. At Level 1 (L1), the lexical level, responses are classified according to their relationship to the target, such as semantic substitutions, perseverations of a previous response, or descriptions of the picture. Phonologically related errors are coded as target attempts (TA), with further specification at Level 2 (L2). Fragment errors are indicated by appending *-f* to the L1 code. At L2, the phonological level, target attempts are classified according to their outcome, most importantly, whether they constitute a sound-related word error (S/W) or a sound-related non-word error (S/NW). Other types of errors may also be coded at L2 if they include a sound-related error. For example, the error *nail* > /hæməl/ would be coded as a semantic error (S) at L1, assuming the substitution of *nail* > *hammer*, and a sound-related error with a non-word outcome (S/NW) at L2. A complete list of Level 1 and Level 2 codes is provided in Appendix 5-v. To ensure the reliability of the coding, all responses were scored by one coder, who was trained on the PNT coding system, and subsequently checked by the author. Ambiguous codings were resolved through discussion and consultation with the authors of the PNT. Because only phonological errors were analyzed in this study, the only further classification required was to divide errors into whole words and fragments, and into word and non-word outcomes.

Fragment Errors: For both tasks, errors were classified as whole-word errors or fragments. In order to count as an error in the NR task, fragments had to deviate from the target by at least one phoneme. In the PNT task, however, fragments in the initial response received an error coding at L1 regardless of whether or not they deviated from the target. For example, the initial responses /kæn-/ and /kɪn-/ for *candle* would both be coded as TA-f at L1. The distinction would be made at L2: /kæn-/ would receive no L2 code, whereas /kɪn-/ would receive a code of S/I to indicate a sound error with indeterminate lexical status. However, because L2 codes were ignored in the present study, reported initial accuracy scores include correct as well as incorrect fragments. This problem is avoided, however, by using accuracy scores from the first complete response.

Lexical Status: In both tasks, whole-word errors were classified as real words or non-words. Real words were identified with reference to the *Shorter Oxford*

Dictionary (Brown, 1993) but, because dictionaries have very liberal criteria for what consitutes a word, no archaic, obsolete or strictly dialectical variants were included. Proper names were also not counted as real words, although slang words were accepted. All inflected forms of a word were considered when classifying lexical status, even if the response was not intended to be inflected. For example, the response "That's a /mænd/" was given to name the picture of a man; /mænd/ was classified as the real word *manned*, even though a singular noun was clearly intended in this context. The rationale here stems from the fact that the error is not *assumed* to reflect a lexical substitution (although it may be, in some cases), but rather a phonological substitution which, either by chance or through the mechanisms of spreading activation, results in a word error.

Statistical Analyses

The role of phonological neighbourhood variables on error production was analyzed in two types of comparisons, as in the pilot study. In the first, the susceptibility of target items to error was assessed. In the NR task, this was accomplished by comparing target items which were produced in error to target items which were produced correctly; in the PNT, the error rates of individual stimulus items were compared. In the second type of comparison, the impact of phonological neighbourhoods on the outcomes of errors was assessed by comparing the errors that were produced to the target items that were intended.

The Neighbourhood Database: Values for item frequency, neighbourhood density and neighbourhood frequency were obtained from an on-line lexicon of 20,000 words based on *Webster's Pocket Dictionary* (see Luce & Pisoni, 1998; Luce et al., 1990). Item frequencies in this lexicon are homophone frequencies, based on Kučera and Francis's (1967) database. This means that the frequencies for all words with the *same* phonological structure are added together, regardless of their orthographic form or grammatical function. For example, the frequency of /kænən/ is a sum of the frequency counts for *cannon* and *canon* but not *cannons*, and the frequency of /bɔrd/ includes counts for *board* and *bored* but not *bore*. Although it may seem more appropriate to include only noun frequency counts for picture stimuli (which is the count cited in the PNT literature), there is evidence that words are influenced by the frequencies of their homophones (Dell, 1990; Jescheniak & Levelt, 1994). It may be that frequency of occurrence exerts different effects at lemma and lexeme levels, but it is phonological frequency which is relevant here, in order to compare any neighbourhood effects with those found in speech recognition studies.

Neighbourhood density represents the number of lexical entries in the database which are phonologically similar to a given item (i.e. its neighbours), where phonological similarity is defined by adding, subtracting or substituting one phoneme of the stimulus item. Neighbourhood frequency represents the average of the frequency counts of all of an item's neighbours. Because this variable is irrelevant for items which have no neighbours, the neighbourhood frequency analyses reflect, for the most part, a restricted set of items with 'zero-density' items removed.

Target Susceptibility Analysis (NR): To assess the slipability of intended targets in the picture description task, the words which were produced in error were compared to a similar set of 'control' words which were correctly produced by the subjects. The control-word corpus was gathered through a pseudo-random selection procedure: For each error made by a given subject during a given picture's description, a correctly produced word, which matched the error's target on grammatical class and number of syllables, was chosen from the same sample. Given the results from the pilot study (as well as previous studies, e.g Buckingham & Kertesz, 1974; Butterworth, 1979; Garrett, 1975) showing a clear difference in the susceptibility of content and function words to error production, it was desirable to control grammatical word class. Furthermore, the knowledge that neighbourhood variables and word length are inter-correlated (Harley & Bown, 1998; Landauer & Streeter, 1973) indicated controlling this factor as well.

The two sets of targets—error-targets and control-targets—were compared using separate *t*-tests for each of the variables: item frequency, neighbourhood density, and neighbourhood frequency. Although Luce (1986) has developed a formula incorporating all of these variables, this formula represents the probability of identification based on results from spoken word recognition studies. As the importance of each of these variables has not yet been established in speech production research, least of all for aphasic subjects, they will be examined here separately.

Error Outcome Analysis (NR): The impact of the phonological neighbourhood on error outcome was examined by comparing the set of errors produced in picture description to their respective targets. Word and non-word errors were examined separately, based on differential results for these two types of errors in the pilot study, and to help distinguish between influences operating at lexical and sub-lexical levels (Vitevitch & Luce, 1998; Vitevitch et al., 1999). An ANOVA with two binary factors—item set (errors *vs* targets) and lexical status of error (words *vs* non-words)—was conducted for the two neighbourhood variables, neighbourhood density and neighbourhood frequency, and for the item characteristic of length, in number of syllables. (Note: this comparison was not relevant in the target susceptibility analysis, because it was controlled through the matching procedure.) Because non-words have no frequency values, the word/non-word error comparison was not possible on this variable. Therefore, two *t*-tests were used to assess item frequency differences between word errors and their targets, and between the targets of word and non-word errors.

Target Susceptibility Analysis (PNT): The issue of target susceptibility was analyzed differently in the PNT task than in the NR task. Because accuracy data were available for each stimulus across the group of subjects, a multiple regression was performed correlating the proportion of errors on each item with its item characteristics (frequency and number of syllables) and its neighbourhood characteristics (neighbourhood density and neighbourhood frequency). By not grouping the data into discrete categories of high and low density, and high and low frequency, as has been done in previous studies (e.g. Luce & Pisoni, 1998; Vitevitch, 1997; Vitevitch, ms in prep), it was hoped that such an analysis would prove more powerful. In addition, a multiple regression allows an assessment of the inter-correlations among the independent variables.

Error Outcome Analysis (PNT): Error outcome was analyzed for the PNT data the same way as for the NR data, using separate 2x2 ANOVAs (item set x lexical status of error) to assess neighbourhood density, neighbourhood frequency, and item length, and *t*-tests to assess item frequency.

Qualitative Analyses

In addition to descriptive analyses of the error corpora, the performance patterns of individual subjects were examined. Individual patterns were compared to the aphasic group as a whole, and to the performance of control subjects (on the NR task). Individual patterns were also analyzed with reference to the subjects' aphasia profiles, in particular, the severity level of aphasia. However, because only one type of error was included in the main study, unlike the pilot study, there was less individual data available for subject comparisons. Furthermore, complete stroke history information to provide independent corroboration of clinical aphasic sub-type classifications was not available for all subjects.

Results

The Error Corpora

As described in the previous section, errors recorded from the transcripts were assessed for their reliability before defining the corpora of errors for each task on which the statistical analysis would be based.

Reliability Assessment

The two independent transcribers recorded a total of 1015 potential phonological errors from the Norman Rockwell picture description task, and 1183 from the *Philadelphia Naming Test*. Through the procedure described in the previous section, all items which were determined by consensus not to meet the criteria for phonologically related errors were excluded. As mentioned, items excluded consisted of errors that were agreed to be correct or normal co-articulations (e.g. *pumpkin* > /pʌŋkən/; *stethoscope* > /stɛtəskop/), variants related to accent or dialect (e.g. *car* > /kɑ/; *zebra* > /zɛbrə/; *chimney* > /tʃɪmbli/), and articulatory distortions and reductions (e.g. *chair* > /tsɛr/; *floor* > /for/). Voicing substitutions were treated especially conservatively, given findings of high rates of phonemic false evaluation on these types of errors (Blumstein et al., 1980, cited in Buckingham & Yule, 1987), and the susceptibility of voicing to articulatory disruption in non-fluent aphasia (e.g. Nespoulous et al., 1984; Trost & Canter, 1974). For some individual subjects, idiosyncratic patterns were noted in the phonological errors recorded. For example, one subject made a large number of /w/ > /r/ substitutions, another added /m/ at the beginning of several words, another fronted velar stops (/k/ > /t/). A subject who was of Scottish origin used vowels different from standard North American pronunciations (e.g. mail > /mil/; beard > /bɜrd/). For these subjects, items were also considered with extra caution in determining whether or not they constituted phonological errors. In all cases, consensus decisions were made with reference to the individual subject's characteristic speech pattern.

Also excluded, as described in the methods section, were errors for which the target could not be unambiguously determined, fragment errors which were too short to be counted, and errors which were determined not to be phonologically related to the target (either unrelated or semantically related). In all, 406 of the 1015 NR items (40.0%), and 385 (32.5%) of the 1183 PNT items were excluded. The initial error lists were extremely liberal in their inclusion; all potential phonologically related errors were recorded by the first two transcribers, so as not to falsely exclude items before the reliability of the transcriptions was assessed. In addition, some items of interest were initially included even though they were not to be analyzed in the current study (e.g. unrelated errors). Thus, because the initial corpus was somewhat inflated, the overall proportions of exclusions should not be taken as a reflection of the reliability of error coding. Nevertheless, it is of interest that the relative proportion of exclusions is greater for the picture description than the naming task, which would be expected for a task which is less structured and for which the targets are not determined *a priori*.

The reliability of the two independent transcriptions (between T1 and T2) also differed across the two tasks: 47.3% for the NR task, and 57.6% for the PNT[1], indicating somewhat greater difficulty with the transcription of running speech. However, T3 was able to resolve the vast majority of these discrepancies. In some cases, a consensus was possible through a 'compromise' transcription, as long as two transcribers agreed on every constituent phoneme. For example, where T1 transcribed /stɛkstɛptop/ for *stethoscope*, T2 transcribed /stɛksəsop/, and T3 transcribed /stɛkstəstop/. All of the transcriptions are different, but T3's transcription agrees with T1's in part, and with T2's in part, so it may be used as the consensus error. For a small proportion of the items in each task—7.0% for the NR task and 4.9% for the PNT—no consensus could be reached among the three transcribers; these items were also excluded. Again, it is notable that consensus was more difficult to achieve on the picture description task than on the naming task. The total number of items ultimately retained in the statistical analyses differs depending on the comparisons being made; the details will be provided as each analysis is discussed.

<u>Norman Rockwell Picture Description Task (NR)</u>

The number of words produced for each of the ten pictures was recorded for each subject. All verbalizations were counted, including errors, fragments, and fillers such as "um" and "oohh", but non-verbal vocalizations (e.g. coughing, laughing, sighing) were not. Untranscribable portions were excluded. Task-related comments were not included (e.g. "Oh, I'm not doing very well"; "Should I continue?"), but personal

[1] Although these reliablility scores are quite low, it appears that many of the discrepancies were due to differences in level of transcription expertise between the transcribers. In general, low reliability at this stage is not surprising for phonetic transcriptions of aphasic errors (c.f. Christman, 1994), and is the motivation behind the extensive reliability procedure.

asides arising from the content of the picture were (e.g. "I have one of those downstairs"; "He's a policeman, just like my brother"). Output cued by the examiner was excluded. Compound words were counted as separate words to avoid making decisions about hyphenation and compounding for each individual word. (For example, are *pig-tail*, *bow tie*, and *cheerleader* one word, two words, or hyphenated?) In addition, compound words are not represented in the neighbourhood database used, so neighbourhood values were only available for each component word. There were exceptions to the rule of splitting compounds, for the most part consisting of words which function as adverbs or prepositions, and which almost always appear as a single, unhyphenated word in current usage (e.g. *anywhere, maybe, inside, outright, himself, otherwise, meanwhile, spoonful*). A list of these words was kept to ensure consistency in the word counts across the different pictures and the different subjects.

Mean word counts for each picture are presented in the first graph in Figure 5-i (p. F-iv) for the six control subjects and the 34 aphasic subjects who successfully completed the task. The two groups show similar patterns in the average numbers of words produced for each picture—that is, certain pictures elicited more output than other pictures for both groups—although the control group consistently produced more words than the aphasic group to describe each picture. The incidence of phonological error production is shown for aphasic subjects in the second graph in Figure 5-i. It is evident that the mean number of errors per picture closely parallels the mean number of words per picture, illustrating that phonological error incidence is largely related to the opportunity for their occurrence, given that task conditions are equivalent across the pictures. It should also be noted that the incidence of phonological errors for the aphasic subjects is very low, ranging from 0.8% to 1.2%, with a mean incidence of 1.0% across the stimulus pictures.

Philadelphia Naming Test (PNT)

Counting all types of errors, overall accuracy scores were obtained for each subject and each item at three levels—the first response, the first complete response, and the final response. As expected (see Roach et al., 1996), the three accuracy levels were highly correlated, as illustrated in the scatter-plots of item accuracy rates in Figure 5-ii (see page F-v). Phonological error rates were also calculated by counting all the phonologically related errors (or 'Target Attempts') which occurred on the first complete response. The correlation between mean overall error rate and phonological error rate on the first complete response for each item is shown in the last scatter-plot in Figure 5-ii. A positive relationship is indicated, with phonological errors occurring more often on items with higher overall error rates. However, the lack of a close correlation suggests that these phonological error rates are not strongly related to item difficulty; their incidence does not increase proportionately with other types of errors for less accurate items. It should be noted, however, that word and non-word outcomes were not differentiated in this measure of phonological error, and that this distinction has been shown to be related to severity of naming impairment (Dell et al., 1997b). Error rates for individual subjects will be presented later.

Statistical Analyses

Target Susceptibility Analysis (NR)

The set of target items consisted of all targets which gave rise to a phonologically related error, including targets of fragment errors. The criterion of phonological relatedness and the context of the utterance, as well as the reliability procedure described above, ensured that the fragments included were indeed attempts at the target identified. In addition, most fragment errors were subsequently self-corrected, providing validation of the intended target. Of the 523 targets produced in error, only 89 (17%) were fragments, and all but 3 of these were immediately self-corrected. If multiple error attempts were made for the same target, the target was counted only once, even if the errors were different. However, a target might recur as a separate attempt by the same subject, and targets often did recur across different subjects. A few of the 'no-consensus' items (less than 2% of the total) were retained for this analysis because, even though the identity of some of the phonemes remained ambiguous, they were nevertheless unambiguously determined to be errors. For example, the target *chair* was transcribed as /tskɛr/ by one transcriber, and *it's* /kɛr/ by another. Even though it is not clear whether the /ts/ was part of the error, there is agreement on the fact of an initial phoneme substitution. The 523 error-targets were matched with 523 control-targets.

Control words were chosen consecutively from the beginning of the sample but, in order to avoid task artifacts, phrases related to the task rather than the content of the picture, such as "I think this is a picture of..." or "In this one we see..." were skipped. As in the error corpus, repeated perseverations and stereotypies were included only once. If no words matching the error's target on both grammatical class and length could be found, the sample for the following picture was used, and so on consecutively through the ten picture samples. If no matching word could be found in any of the samples for a given subject, as was occasionally the case for multisyllabic words, either a length-matched word from a different grammatical class, or a grammatically matched word which approximated the target's length as closely as possible, was selected. Of the 523 target/control pairs, 17 (3.3%) did not match on grammatical class, but in all cases the content/function word distinction was maintained. Fifteen pairs (2.9%) did not match on length; 12 of these differed by one syllable and 3 pairs differed by two syllables. A list of the matched error-target and control-target words can be found in Appendix 5-vi.

The error-targets and control-targets were submitted to the on-line lexicon in orthographic form to obtain values for their item frequency, neighbourhood density and neighbourhood frequency. The database includes only uninflected forms, so targets were first stripped of their inflectional morphemes (e.g. number and tense markings). Nevertheless, because the database is small relative to the entire vocabulary, some words were not found in the lexicon. In these cases, items were submitted in phonetic form to obtain neighbourhood values, and item frequencies were calculated from Francis and Kuèera (1982), using the same procedures as were used for the database (i.e. calculating phonological (or homophone) frequencies). Frequency values, which often tend to be skewed towards the higher frequency end of the distribution, are log-transformed in many studies to solve this problem (e.g. Goldinger et al., 1989; Vitevitch, 1997). The log values for item and neighbourhood frequency, which are provided by the database, were also used in the current study. After looking at the distribution of density values, the decision was made to log-transform this variable as well. Furthermore, because neighbourhood frequency is irrelevant for zero-density items, the neighbourhood frequency analysis included only items with a density of one or more.

Raw means for the error-target and control-target corpora are compared graphically in Figure 5-iii (see page F-vi). The control corpus shows higher values of all three variables—higher mean item frequency, greater mean density, and higher mean neighbourhood frequency. Log values were analyzed on each neighbourhood variable using separate *t*-tests. A significant effect of log frequency was found, such that the targets of the errors were of lower frequency overall than the control words ($t_{(1044)}$ = -

6.98, $p < 0.001$). Error-targets were also found to come from significantly less dense neighbourhoods than the control words ($t_{(1044)} = -2.58$, $p < 0.05$). No significant effect of log neighbourhood frequency was found ($t_{(903)} = -1.531$, $p = 0.126$). Thus, the items which were produced in error in this task represent a less frequent set of words with sparser neighbourhoods than a comparable set of control words from the same samples. Item frequency and neighbourhood density appear to influence the susceptibility of targets, even when grammatical class and word length are controlled.

Error Outcome Analysis (NR)

The set of errors produced during the Norman Rockwell picture description task was compared to the set of corresponding targets. Only whole-word errors were included in this analysis but, unlike the susceptibility analysis, multiple attempts at each target were included (as long as they were all whole-word attempts). As in the previous analysis, it was necessary to remove inflections from target items before submitting them to the on-line lexicon. To maintain comparability between the targets and their errors, inflections were removed from the errors as well, as long as it was clear what the inflections were. For example, in the error /wɛðərɪn/ produced for the target *wearing*, /wɛðər/ was submitted for the error, and *wear* for the target. If, however, the error appeared to be inflected, but the inflection was inappropriate to the context (as in the example given earlier of "That's a /mænd/" produced for *man*), it could not be confidently determined that the inflection was intended, so the error was submitted to the database as a whole. Word errors were submitted in orthographic form; non-word errors in phonetic form.

Figure 5-iv (p. F-vii) shows the raw mean values for error and target corpora in the NR task on the item measures of frequency and length, and on the neighbourhood

measures of density and neighbourhood frequency. On the whole, word errors and their targets appear to differ considerably from non-word errors and their targets, although the error/target differences, for the most part, appear negligible. One glaring exception is the apparently huge difference in raw mean frequency between word errors and their targets, with the errors being less frequent than their targets, an unexpected finding. The size of the difference may be due to a few extremely high-frequency items in the target set, but the log-transformation of the values should minimize their impact in the statistical analyses. It is also interesting that the density differences between errors and targets are in opposite directions for word and non-word errors. The groups of items were analyzed statistically, as described in the previous section, by *t*-test for item frequency and by ANOVA for neighbourhood density, neighbourhood frequency, and word length.

Frequency: Using log-transformed frequency values, word errors were compared to their targets, and the targets of word errors were compared to the targets of non-word errors. A highly significant effect of lexical status was found for the targets, such that the targets of word errors were higher in frequency than the targets giving rise to non-word errors ($t_{(448)} = 4.72$, $p < 0.001$). In addition, a strong effect of item set was shown ($t_{(434)} = -4.56$, $p < 0.001$), whereby word errors were significantly less frequent than their targets. Thus, despite log-transforming the data, the frequency difference observed in the raw means was still statistically significant. The possible reasons for this anomalous result will be discussed in the next chapter.

Density: Two factors were included in the ANOVA, each with two levels: 'item set' divides errors from targets, irrespective of lexical status; 'lexical status of error' distinguishes word errors together with their targets from non-word errors and their targets. Density values were again transformed logarithmically. A strong main effect of lexical status was shown, with the word error set having significantly denser neighbourhoods than the non-word error set ($F_{(1, 896)} = 291.96$, $p < 0.001$). There was no significant main effect of item set ($F_{(1, 896)} = 1.07$, $p = 0.302$); however, the interaction between item set and lexical status of error was significant ($F_{(1, 896)} = 5.37$, $p < 0.05$). Post-hoc analyses were conducted to explore this interaction. Using Newman-Keuls comparisons, word errors were not found to be significantly different from their targets, but non-word errors were shown to come from significantly less dense neighbourhoods than their targets ($p < 0.05$). Word errors were from significantly more dense neighbourhoods than non-word errors ($p < 0.01$); similarly, targets of word errors were from significantly more dense neighbourhoods than targets of non-word errors ($p < 0.01$). These findings are not surprising given the strong main effect of lexical status.

Neighbourhood Frequency: The ANOVA conducted on neighbourhood frequencies had the same design as that for neighbourhood density, except that all items with a density value of zero were first eliminated from the data sets. This step was intended to improve the distribution of items on this variable by removing those items for which neighbourhood frequency is irrelevant (because there are no neighbours for which to calculate frequencies). Of the original 900 items, 145 zero-density items (16%) were removed. Using log neighbourhood frequency values, a significant main effect of the lexical status of the error was revealed, such that the word-error set had a higher neighbourhood frequency value than the non-word error set ($F_{(1, 751)} = 21.06$, $p < 0.001$). A significant main effect of item set was also found, with errors showing significantly higher neighbourhood frequencies than targets ($F_{(1, 751)} = 12.56$, $p < 0.001$). The interaction between item set and lexical status of error was not significant ($F_{(1, 751)} = 2.17$, $p = 0.145$).

Word Length: A 2x2 ANOVA, as described above, was conducted on the raw values for number of syllables in each item set. Again, a significant main effect of lexical status was shown, with word error set having significantly fewer syllables than non-word error set ($F_{(1, 896)} = 185.24$, $p < 0.001$). The main effect of item set, however, was not significant ($F_{(1, 896)} = 0.37$, $p = 0.54$), and neither was the interaction ($F_{(1, 896)} = 0.37$, $p = 0.54$).

Target Susceptibility Analysis (PNT)

The role of item and neighbourhood variables in predicting the susceptibility of the PNT naming stimuli to error was assessed using a backward step-wise multiple regression analysis. As a preliminary step, each variable was correlated separately with the error rates of the first complete responses, averaged across subjects for each naming stimulus. The correlation scatter-plots and their respective r-values are presented in Figure 5-v.a (p. F-viii). The log-transformed values for item frequency and neighbourhood density show moderate negative correlations with the log-transformed error rate, suggesting that error rates increase as item frequency and neighbourhood density decrease. Word length also shows a moderate, but positive, correlation with error rate, illustrating that errors, not surprisingly, are more frequent on words with more syllables. The log-transformed values of neighbourhood frequency (excluding those items with zero density) show a very small negative correlation with error rate.

In Figure 5-v.b (p. F-ix), the inter-correlations among the item and neighbourhood variables are illustrated. As anticipated, a moderate correlation was found between density and item frequency, and a moderate-to-high correlation between density and neighbourhood frequency. The more frequent an item, the more neighbours it has; the more neighbours an item has, the higher its neighbourhood frequency. Word length is also highly negatively correlated with density, as expected, and, because density and neighbourhood frequency are inter-correlated, it is not surprising to find that word length is also moderately negatively correlated with neighbourhood frequency. Thus, the more syllables in a word, the fewer neighbours it has, and the lower its neighbourhood frequency. Small-to-moderate correlations are noted between word length and item frequency, and between item frequency and neighbourhood frequency. (Again note that all neighbourhood frequency correlations contain a restricted set of items with zero-density items removed.)

The aim of the regression analysis was to reveal which of these variables appear to be the most important predictors of naming accuracy. Because of the missing neighbourhood frequency values for zero-density items, this variable was excluded so that the regression analysis could be conducted on the full set of PNT stimuli. The correlations described above suggest that the exclusion of neighbourhood frequency would not sacrifice much predictive power from the regression model. Thus, the log-transformed error rates were regressed on the log values of item frequency and neighbourhood density, and on the number of syllables of each item. In the original model, the three variables together accounted for 26.0% of the variance in error rates ($R = 0.510$). In the next step, the variable contributing the least amount to the model—number of syllables—was removed, with negligible change in the overall R (from 0.510 to 0.507). Both the remaining variables—log frequency and log density—showed significant contributions to the model ($p < 0.001$), so no further steps were taken. In the resulting model, the two variables together made a significant contribution in accounting for the error rates ($F_{(2, 172)} = 29.72$, $p < 0.001$); however, it should be noted that they still only accounted for about 26% of the variance in PNT accuracy.

Error Outcome Analysis (PNT)

As for the Norman Rockwell task, errors were compared to their targets to determine the effect of neighbourhood variables on error outcome, and the two item sets were split into word errors and non-word errors. The set of errors for this analysis, as in the NR task, consisted of whole-word errors only, but did include multiple (whole-word) attempts at the same target. The items differed somewhat from those used in the PNT regression analysis, in that errors made on the practice items were included, as well as phonological errors on 'alternate targets', such as semantic substitutions. For example, the error /vailəlain/ was produced for *guitar*. In the target susceptibility analysis, this error simply registers as an error on the target *guitar*, whereas in the error outcome analysis, /vailəlain/ is entered as the error, and *violin* as the target.

In Figure 5-vi (p. F-x), untransformed mean values for item frequency, length, neighbourhood density and neighbourhood frequency are shown for errors and targets in the word and non-word error sets. The pattern of means in the different sets of items is very similar to that shown in the Norman Rockwell task, with evident differences between word and non-word sets, and negligible differences between errors and targets, except for an apparent item frequency difference between word errors and their targets. This difference is in the expected direction, with word errors being higher in frequency than their targets. Also of interest are the apparent interactions for neighbourhood density and neighbourhood frequency; in both graphs, it can be seen that differences between error and target sets are in opposite directions for word and non-word error sets. As in the NR error outcome analysis, *t*-tests were used to assess the effects of item frequency, whereas ANOVAs were used to assess the effects of both neighbourhood variables—neighbourhood density and neighbourhood frequency—and the item variable word length. All variables except word length were log-transformed.

Frequency: A significant effect of lexical status was shown for the targets, such that the targets of word errors were of higher frequency than the targets of non-word errors ($t_{(588)} = 6.90$, $p < 0.001$). Unlike the picture description task, no effect of item set was found; word errors were not significantly different in frequency from their targets ($t_{(572)} = -0.84$, $p = 0.40$), despite the large difference in raw means. It appears that, as observed for the NR task, the differential between the raw means reflects a few very high-frequency items in the word error set, but that the log-transformation minimized the impact of these outliers.

Density: The density effects observed were similar to the NR picture description task. A significant main effect of lexical status was shown ($F_{(1, 1176)} = 537.15$, $p < 0.001$), with word errors and their targets coming from denser neighbourhoods than non-word errors and their targets. The main effect of item set (errors *vs* targets) was not significant ($F_{(1, 1176)} = 1.59$, $p = 0.21$), but there was a significant interaction between the two factors ($F_{(1, 1176)} = 8.43$, $p < 0.005$). Post-hoc analyses using Newman-Keuls pair-wise comparisons showed the expected significant effects of lexical status of error for both errors and targets ($p < 0.01$), corresponding to the direction of the main effect. In addition, a significant effect of item set was found for the non-word set, with non-word errors having less dense neighbourhoods than their targets ($p < 0.01$). No item set effect was shown for words.

Neighbourhood Frequency: As in the picture description task, items with zero density were excluded before running the analysis. Of the total 1180 items, 252 (21.3%) were eliminated. With this revised set of items, highly significant main effects were found for both lexical status of error ($F_{(1, 924)} = 77.07$, $p < 0.005$), and item set ($F_{(1,}$

$_{924)} = 9.47$, $p < 0.001$): word errors and their targets had higher neighbourhood frequencies than non-word errors and their targets, and errors had higher neighbourhood frequencies than targets. The interaction between the two factors was not significant ($F_{(1, 924)} = 1.36$, $p = 0.243$), even though the raw means indicate that the differences between errors and targets are in opposite directions for word and non-word errors. Again, the lack of effect is likely related to the transformation of the data.

Word Length: In the ANOVA conducted on number of syllables, a highly significant main effect of lexical status was again shown ($F_{(1, 1176)} = 367.24$, $p < 0.001$), where items from the word error set had fewer syllables than items from the non-word error set. The main effect of item set (errors *vs* targets) was not significant ($F_{(1, 1176)} = 0.48$, $p = 0.49$); nor was the interaction between the two factors ($F_{(1, 1176)} = 1.76$, $p = 0.18$).

Task Comparisons

Similar results were found in the NR and PNT tasks, despite differences in the nature of the two tasks. For example, arising from connected discourse, the targets in the picture description task may include words from any grammatical class, whereas the targets on the naming task include only pictureable nouns. Although the PNT was developed to include items with a wide range of frequencies, and a range of word lengths (implying also a range of densities and neighbourhood frequencies), the targets represent a considerably more restricted group of targets than the potential targets in the NR task. The distributions of item frequency, density, and neighbourhood frequency for targets from both tasks were plotted to see whether any obvious differences could be observed. Because there was no predefined set of targets for the picture description task, the targets of the errors were plotted, as well as the control-targets. Figure 5-vii shows the distributions of the three sets of target items on the variables of interest: in Figure 5-vii.a (p. F-xi), the item variables (item frequency and word length) are presented; in Figure 5-vii.b (p. F-xii), the neighbourhood variables (density and neighbourhood frequency) are graphed. Raw values are presented for the density and length distributions, as these are more meaningful; however, because the frequency distributions are so large, it was necessary to present the log-transformed values for item and neighbourhood frequency. For each variable except word length, which is already categorical, the items are divided into interval categories for ease of presentation.

It is clear that the distributions are very similar across the three sets of targets. For item frequency, the distribution of PNT targets is shifted somewhat towards the lower frequency end of the continuum compared to the NR distributions, probably reflecting the fact that the stimuli consist only of content words. The distribution of control-targets in the NR task is shifted towards the higher frequency end, relative to the error-targets, which is reflected in the statistical comparisons, but the overall shapes of the two distributions are very similar. The distributions of word length are almost completely overlapping in the three groups of targets, and the distributions of neighbourhood variables are also very close. The similarities across the three distributions validate the set of PNT stimuli as fairly representative of naturally occurring words (content words, at least) in its item and neighbourhood characteristics. Note that the distribution of density here (0 to 40 neighbours) is almost 25% broader than the distribution of density of the items used in the pilot study (0 to 31 neighbours); thus, the lack of significant density effects for naming in that experiment may well be explained by the limited range in densities in the stimulus set. The neighbourhood frequency graph in Figure 5-vii.b shows the complete set of stimuli, and the over-representation of 'zero' items (reflecting 'zero-density' items) is evident.

Target Susceptibility

A summary of the statistical results of the target susceptibility analyses in each task is presented in Table 5-i (following). Results of the two tasks are largely convergent. Both tasks illustrate that the frequency of occurrence of a target, as well as its neighbourhood density, play an important role in determining the accuracy of its production. The moderate correlation of word length with PNT naming accuracy suggests that this variable also affects production accuracy. However, the regression analysis indicates that its role is largely redundant with the other two variables; this is supported by the moderately high correlation ($r = -0.76$) shown between word length and neighbourhood density (see Figure 5-v.b). Furthermore, the effect of density in the NR task was significant, even though word length was controlled. The effect of neighbourhood frequency was not significant in the NR task and, although it was not included in the multiple regression of PNT accuracy, its low simple correlation with PNT error rate ($r = -0.18$) is consistent with the null effect found in the NR task.

Table 5-i. Target Susceptibility Effects in NR and PNT Tasks

Comparison	Statistical Test	Probability Value
NR Task		
Item Frequency	$t = -6.98$	$p < 0.001$
Neighbourhood Density	$t = -2.58$	$p < 0.050$
Neighbourhood Frequency	$t = -1.53$	$p = 0.126$
PNT Task		
Individual Correlations		
Item Frequency	$r = -0.44$	
Neighbourhood Density	$r = -0.41$	
Neighbourhood Frequency	$r = -0.18$	
Word Length	$r = +0.40$	
Regression Model		
(Item Freq. & N. Density)	$R = -0.51$; $F = 29.72$	$p < 0.001$

Despite these significant findings, it should be noted that the amount of variance accounted for by item frequency and neighbourhood density in the PNT task is small

(26%). It seems logical that this reflects the influence of other factors on naming accuracy, since the error rate used in the analysis includes other types of errors, such as semantic errors, descriptions, and no responses. However, when the regression analysis was re-run using phonological error rates, an even smaller amount of the variance was accounted for (20%). Perhaps this is due to the fact that even the phonological error rate does not factor out all other influences; for example, semantic effects in the mixed errors and contextual effects in perseveratory errors may over-shadow the effects of the phonological variables. On the other hand, it may simply be a statistical artifact of the very low rates of phonological errors produced in this task.

Error Outcome

The results of the error/target comparisons in the two tasks are presented in Table 5-ii (see following page). Again, results are very similar in the NR and PNT tasks: significant effects of the lexical status of error (words *vs* non-words) are shown for all the variables, whereas the effect of item set (errors *vs* targets) is significant in only five of the twelve comparisons. This relative lack of item set differences suggests that phonological errors, despite the liberal way in which they were defined, tend to preserve many of the phonological characteristics of their targets. Comparison of the raw means (see again Figures 5-iv and 5-vi) illustrates that errors almost always consist of the same number of syllables as their targets, and tend to have the same neighbourhood density as well (which, of course, is highly correlated with word length).

Although no main effects were found in either task in the density comparisons of errors and targets, in both tasks there was an interaction between the item set and the lexical status of the error, such that word errors had higher neighbourhood densities than their targets, and non-word errors had lower neighbourhood densities than their targets. These differences were only significant for the non-word error/target

Table 5-ii. Error Outcome Effects in NR and PNT Tasks

Comparison	Statistical Test	Probability Value
NR Task		
Item Frequency		
lexical status of error (targets only)	$t = 4.72$	$p < 0.001$
item set (words only)	$t = -4.56$	$p < 0.001$
Neighbourhood Density		
lexical status of error	$F = 291.96$	$p < 0.001$
item set	$F = 1.07$	$p = 0.302$
interaction	$F = 5.37$	$p < 0.05$
post-hoc tests		
W vs NW errors		$p < 0.01$
W vs NW targets		$p < 0.01$
W errors vs targets		n.s.
NW errors vs targets		$p < 0.05$
Neighbourhood Frequency		
lexical status of error	$F = 21.06$	$p < 0.001$
item set	$F = 12.56$	$p < 0.001$
interaction	$F = 2.17$	$p = 0.145$
Word Length		
lexical status of error	$F = 185.24$	$p < 0.001$
item set	$F = 0.37$	$p = 0.54$
interaction	$F = 0.37$	$p = 0.54$
PNT Task		
Item Frequency		
lexical status of error (targets only)	$t = 6.90$	$p < 0.001$
item set (words only)	$t = -0.84$	$p = 0.40$
Neighbourhood Density		
lexical status of error	$F = 537.15$	$p < 0.001$
item set	$F = 1.59$	$p = 0.21$
interaction	$F = 8.43$	$p < 0.005$
post-hoc tests		
W vs NW errors		$p < 0.01$
W vs NW targets		$p < 0.01$
W errors vs targets		n.s.
NW errors vs targets		$p < 0.01$
Neighbourhood Frequency		
lexical status of error	$F = 77.07$	$p < 0.001$
item set	$F = 9.47$	$p < 0.005$
interaction	$F = 1.36$	$p = 0.243$
Word Length		
lexical status of error	$F = 367.24$	$p < 0.001$
item set	$F = 0.48$	$p = 0.49$
interaction	$F = 1.76$	$p = 0.18$

comparisons, which may reflect the *relative* lack of constraints on non-word phonological structure. The word 'relative' is stressed, because previous research has illustrated that error production in general is highly constrained by numerous language factors, as discussed in the first chapter. This issue is discussed further in the next chapter. Even though no significant density differences were found overall between errors and targets, and the simple effects found between non-word errors and their targets were in the direction opposite to that predicted, nevertheless a significant difference in neighbourhood frequency was found between errors and targets in both tasks. Thus, errors came from higher frequency neighbourhoods than targets, as predicted, and this effect cannot be attributed to the positive correlation between neighbourhood density and neighbourhood frequency.

A significant difference between errors and targets was also noted for item frequency (words only), but in the NR task only. Furthermore, this difference was in the opposite direction to what was predicted, with word errors significantly less frequent than their targets. It is possible that this result is due to a few very high-frequency items amongst the set of targets in the NR task, such as the word '*two*' whose frequency value adopts the frequencies of the function words '*to*' and '*too*', in addition to its own frequency count. Even though these items only occurred a few times, their phonological frequency values are exponentially larger than the rest of the set, a disparity for which the log transformation was perhaps unable to compensate. Nevertheless, because the errors and targets represented here were drawn from a natural sample of speech, it was of theoretical interest to look at all the data, without removing any of the data points.

Subject Analyses

In Table 5-iii (following page), the subjects who participated in both of the experimental tasks are listed along with their performance measures on the BDAE (severity level and average percentile scores for the auditory comprehension and naming sub-tests), the NR task (mean number of words produced per picture, mean proportion of phonological errors per picture), and the PNT (overall error rate and phonological error rate on the first complete response). Subjects are ranked first by severity level (from least severe—5, to most severe—1), and within each severity level by auditory comprehension then naming percentiles, to illustrate the relationship between severity of aphasia and error production. These relationships are displayed graphically in Figure 5-viii (p. F-xiii).

Figure 5-viii compares the distributions of error rates to severity level for each subject. In the first graph, subjects are ranked by the proportion of errors made on the first complete responses of the PNT, from most to least errors. This overall rate is inversely related to the BDAE severity level of the subjects, as indicated by the divergence of the two profiles. In the second graph, subjects are ranked by the proportion of phonological errors made on the first complete response of the PNT, but the proportion of phonological errors on the NR is also plotted. There is much less correspondence between phonological error rates and BDAE severity level than between overall error rate and severity. Furthermore, it is apparent that the ranking of phonological error rates across subjects differs somewhat between the two tasks; the subjects who made the most phonological errors in the PNT are not the same subjects who made the most phonological errors in the NR task. Thus, it appears that, in this group of subjects, phonological errors are not in themselves an indication of aphasia

Table 5-iii. Subject Performance Across Tasks and Screening Tests

		BDAE scores		NR scores		PNT scores	
Subject	Sev.[1]	AC %ile	Naming %ile	No. of words	%Phon Errors	%All Errors	%Phon. Errors
M08	5	95	100	178.3	0.1%	4.0%	0.0%
YD11	5	95	99	228.8	0.7%	6.3%	1.1%
YD04	5	94	99	142.2	0.1%	4.6%	0.6%
M19	5	89	98	239.2	0.3%	10.3%	2.3%
M25	4	91	97	55.8	3.6%	19.2%	7.0%
YD14	4	91	93	184.7	2.9%	14.3%	10.3%
M12	4	90	98	123.8	0.9%	5.1%	2.3%
YD10	4	90	96	455.9	0.2%	9.7%	0.0%
M22	4	89	99	227.0	3.7%	17.1%	14.3%
M13	4	89	94	330.8	0.2%	17.2%	1.7%
M14	4	88	94	360.7	1.0%	15.4%	2.3%
M01	4	86	98	166.0	2.3%	22.9%	13.7%
YD07	4	86	96	158.3	1.3%	25.7%	13.7%
YD16	4	86	95	403.8	0.5%	17.1%	5.1%
M20	4	82	94	68.8	0.0%	21.1%	2.3%
M17	4	81	90	135.4	1.3%	17.7%	0.0%
YD02	4	80	92	110.9	0.2%	24.6%	2.3%
YD17	4	74	98	75.4	0.7%	11.4%	0.6%
M16	3	84	83	80.2	3.2%	33.9%	2.9%
M10	3	79	74	107.6	0.5%	44.8%	3.4%
YD12	3	78	78	95.1	0.7%	26.9%	0.6%
M18	3	77	50	102.2	0.3%	34.5%	1.1%
YD06	3	74	90	315.6	0.5%	20.0%	1.1%
M15	3	72	95	119.4	0.2%	33.5%	2.3%
YD15	3	71	80	82.6	3.8%	41.6%	13.9%
M24	3	65	90	171.1	1.9%	41.1%	13.1%
YD05	2	66	35	233.4	0.6%	96.0%	6.9%
M21	2	60	70	122.8	2.5%	39.3%	13.3%
YD08	2	56	67	41.5	5.1%	33.1%	12.6%
YD09	2	39	63	68.8	2.2%	34.3%	17.1%
Mean	3.6	79.9	86.8	172.9	1.4%	24.8%	5.6%
Max	5.0	95.0	100.0	455.9	5.1%	96.0%	17.1%
Min	2.0	39.0	35.0	41.5	0.0%	4.0%	0.0%

[1]severity ranking (5=least severe; 1=most severe)

severity, a finding which conflicts with some previous reports in the aphasia literature (e.g. Dell et al., 1997b; Mitchum et al., 1990; but see Moerman et al., 1983).

In the pilot study, subjects' clinical sub-type (i.e. fluent *vs* non-fluent) was related to the proportions of phonological and non-phonological errors, and to the relative proportions of different types of contextual errors. However, these analyses were not the goal of the main study, which focused instead on phonologically related errors only. Because only one type of error was analyzed here, a detailed analysis of error patterns across individual subjects was not possible.

Chapter 6. Neighbourhoods in Aphasic Speech Production

The goal of the current study was to investigate the retrieval of phonological word forms during aphasic speech production, in order to inform models of the structure and function of the phonological lexicon. Using a naturalistic, less structured task—picture description—and a more structured, single-word production task—picture naming—several characteristics of the target and its phonological neighbourhood were examined, specifically: the target word's frequency of occurrence; the number of words which are phonologically similar to the target, or its neighbourhood density; and the average frequency of those neighbours, or neighbourhood frequency. (The target word's length was also included in the analyses, to factor out its contribution to the effects observed.)

To assess the influence of these factors on a target's susceptibility to error, the frequency, neighbourhood density, and neighbourhood frequency values of the words produced incorrectly in the picture description task were compared to those values of a comparable corpus of correctly produced words from the same speech samples. In the naming task, target susceptibility was assessed by analyzing the error rates on individual stimulus items, as a function of their frequency, length, and neighbourhood values. The results of both tasks indicated that the lower a target's frequency of occurrence was, and the fewer neighbours it had, the more susceptible it was to error. Neighbourhood frequency, however, did not appear to have an impact on target susceptibility.

To assess the impact of the neighbourhood on error outcome, the item frequency, length, neighbourhood density, and neighbourhood frequency values of the errors produced were compared to those of their targets. Word and non-word errors were analyzed separately. In neither task were errors found to differ consistently from their targets on the variables of item frequency and neighbourhood density. Errors were, however, found to have higher neighbourhood frequencies than their targets, suggesting that a higher frequency neighbourhood promotes accurate production. Consistent differences were also found on all the variables between word error/target and non-word error/target sets.

These results contribute to the literature on lexical access primarily by extending findings of neighbourhood effects in normal speech production to the aphasic population. In doing so, the present study lends support to the basic tenets of the Neighborhood Activation Model (Luce & Pisoni, 1998), and to the notion of the continuity thesis (Dell et al., 1997b). Although the aim of the study was not to test predictions of interactive *vs* serial models, results are interpreted as consistent with an interactive connectionist framework of speech production, to be discussed shortly.

While not the focus of this study, results also corroborate some of the findings from the aphasic speech-error literature. In the analysis of contextual errors in the pilot study, for example, support was shown for the finding that aphasic subjects tend to exhibit a larger proportion of perseverative than anticipatory errors (e.g. Dell et al., 1997a; Schwartz et al., 1994), unlike normal subjects (e.g. Dell et al., 1997a; Garnham et al., 1981). Some of the results from the main study also have precedents in the literature, such as the impact on target susceptibility of lexical frequency (Blanken, 1990; Ellis et al., 1983; Gagnon & Schwartz, 1996) and word length (Best, 1996; Favreau et al., 1990; Friedman & Kohn, 1990; Kohn & Smith, 1994a; Nickels & Howard, 1995; Pate et al., 1987; Romani & Calabrese, 1998). Furthermore, preliminary speculations about the importance of phonological neighbourhood density for accurate speech production (e.g. Best, 1995; Dell et al., 1997b; Gagnon et al., 1997; Nickels & Howard, 1995) have been confirmed. In general, the results are in agreement with previous studies illustrating that aphasic error outcomes are strongly constrained by a number of linguistic factors which also constrain normal error production (see Buckingham, 1980; Dell et al., 1997a; Harley & MacAndrew, 1992; Stemberger, 1982b; Talo, 1980). More detailed analyses of the patterns of errors produced (e.g. types and relative proportions of errors, error/target relationships, differences between individual subjects) were not the goal of the present study, and await future investigations.

The most important result of the current study, then, is the finding that both lexical frequency and neighbourhood density exert a facilitative effect on the accurate retrieval of words in aphasic speech production, just as they do in normal speech production (Harley & Bown, 1998; Vitevitch, 1997; Vitevitch, ms in prep; but see Jescheniak & Levelt, 1994). Neighbourhood frequency was not found to have a significant impact on ease of lexical access, but this effect has also been shown only inconsistently in normal studies (e.g. see Luce & Pisoni, 1998; Vitevitch, 1997; Vitevitch, ms in prep, Exp. 1). Neighbourhood variables were not found to significantly influence the outcomes of errors, and this finding also mirrors normal error studies (e.g. Vitevitch, 1997).

The effect of density provides support for the Neighborhood Activation Model (Luce & Pisoni, 1998) in its contention that the number of words phonologically related to a target will influence its retrieval; however, the direction of the effect is contrary to the predictions of the NAM. It has been proposed that both competitive and facilitative effects in word recognition may be accommodated within the NAM at different linguistic levels—competitive effects at the lexical level, and facilitative effects at a sub-lexical level (Vitevitch & Luce, 1998; Vitevitch et al., 1999). However, the results of recent studies suggest that such an account cannot fully explain the facilitative neighbourhood effects found in studies of normal error production (Vitevitch, 1997; Vitevitch, ms in prep). Moreover, unlike receptive tasks involving repetition of non-words (Vitevitch & Luce, 1998) or same-different judgements (Vitevitch & Luce, 1999), there is no reason to expect sub-lexical effects to dominate in tasks such as picture naming and picture description, which require the retrieval of both lemmas and lexemes of real words in relatively natural speech production circumstances. Thus, it is unlikely that a sub-lexical explanation would account for the aphasic error production data.

Target Susceptibility Effects: Interactivity Revisited

Interactive spreading activation models, on the other hand, provide a mechanism which is able to account for the facilitative (or 'conspiracy') effects of phonological neighbourhood factors. For example, Vitevitch (Vitevitch, ms in prep) refers to MacKay's (1987) interactive model of Node Structure Theory; in the present discussion, results are explained within Dell's interactive model of speech production (Dell, 1985, 1986, 1988). Within this framework, the facilitative effects of neighbourhood density and frequency are hypothesized to occur through activation spreading along the excitatory and bidirectional connections between word nodes and phoneme nodes in the phonological lexicon (Dell, 1988). In other words, activation spreads from higher-level units (i.e. words, or lexemes) to their respective lower-level constituents (i.e. syllables and phonemes), then reverberates back to higher-level units, a sort of "mutual backscratching" (Dell, 1985). Figure 6-i (page F-xiv) illustrates this concept. Thus, activation spreads from a target word to its phonological neighbours and back again *via* their shared phonemes, which in turn reinforces the activation of the target (Dell, 1986). The more phonological neighbours a target has, the greater the amount of reinforcement it receives from the phoneme level (Vitevitch, 1997; Vitevitch, ms in prep).

As described in Chapter 3, this proliferation of activation occasionally results in the production of unintended but related items. Errors occur when items related to the target through semantic or phonological connections accidentally achieve higher levels of activation than the target itself: "the background activation of the lexical network provides a source of variability in the patterns of activation that result during production" (Dell, 1986, p. 291). Another source of variability is the 'noise' arising from the activation of items unrelated to the intended message, for example from environmental distractions, intruding thoughts, or perseverated words (see Harley, 1990). At the decision stage, the most highly activated units are selected for production. According to Dell (1986), "the activation pattern among the sound nodes adjusts itself over time so that the most highly activated nodes correspond to a single morpheme or word" (p. 300). However, if no single word is sufficiently activated, the pattern of activation existing among phoneme nodes at the decision stage may result in the production of a non-word error. Once an item has been selected, it is inhibited to prevent excessive spreading of activation, but will continue to receive activation reverberating from connected nodes. This post-selection inhibition operates in much the same way as Shattuck-Hufnagel's (1979) check-off monitor (Dell, 1988).

Although the feedback connections in the model can account for the occurrence of many types of errors, their existence is actually motivated, somewhat paradoxically, by their ability to "edit out potential production errors" (Dell, 1985, p. 7). Because activation reverberates among connected units, the target item is, under normal circumstances, the one most likely to achieve the highest level of activation. Furthermore, the mechanism of activation feedback means that, should an error occur, it is more likely to be a real word which is related to the target, since these items receive more reverberating activation than either unrelated words or patterns of phonemes corresponding to non-words. More frequent phonemes and phoneme sequences also receive a greater amount of activation due to the fact that they are connected to more items within the network. Thus, Dell's model is able to account for findings such as phoneme frequency effects, lexical bias, and phonological facilitation without the construction of separate modules with the sole function of monitoring output. Editing is carried out automatically through the interactive spread of activation. But, Dell notes, "[l]ike an editor, the feedback system only works well if it has enough time... the greater the opportunity for activation to reverberate between morphemes and phonemes, the greater the likelihood of editing out nonmorphemes" (1985, p. 10).

The unintended activation of phonological neighbours may also be explained within non-interactive models, but it is not clear how such models would account for the *facilitative* effects of neighbourhood density and frequency. Levelt and colleagues (Levelt et al., 1999; Levelt et al., 1991a) do not claim that their serial-stage model can account for all patterns of errors, preferring instead to focus on the production of correct language; they state that the "ultimate test" of models of lexical access "cannot lie in how they account for infrequent derailments of the process but rather must lie in how they deal with the normal process itself" (Levelt et al., 1999, p. 2). Nevertheless, phonologically related errors may be explained through the same mechanism of forward spreading activation which is hypothesized to account for priming effects in their reaction time experiments. Not surprisingly, this mechanism differs from Dell's in the absence of feedback connections. By contrast, the selection of the target over its competitors is assured through a 'binding' mechanism similar to Shattuck-Hufnagel's scan-copier (1979), which checks the correspondence of lower-level nodes with their attached nodes one level higher. (Contextually related phonological errors are accounted for through occasional failures of this device.) As an alternative to feedback, Dell (1985) suggested that phonologically related words might be activated through lateral excitatory connections among word forms. However, he concluded that such a mechanism does not account for the time-dependency of effects such as lexical bias. Furthermore, any mechanism for activating neighbours must rely on phonological similarity, and therefore most logically operates through the phoneme nodes. Dell states: "If it must act like feedback why not just admit that it is feedback?" (1985, p. 20), a criticism which applies equally well to Levelt's (Levelt et al., 1999) binding mechanism.

Error Outcome Effects: Preservation of Constraints

Effects of Item Set

The absence in both tasks of neighbourhood density and word length effects on error outcome indicates that errors were phonologically similar to their targets. On average, errors contained the same number of syllables as their targets, and had approximately the same number of neighbours. That density was systematically maintained in errors implies that errors frequently came from the target's neighbourhood, or a nearby neighbourhood. (Of course, it is possible for words with entirely different phonological shapes to have similar density counts, but such coincidental effects are not likely to occur with statistical reliability.) Because density and length are closely related (Best, 1995; Harley & Bown, 1998; Landauer & Streeter, 1973), it is not surprising that null effects were found for both variables; however, the density constraint indirectly implicates a number of other phonological constraints as well, such as the preservation of syllabic structure and stress contours. Thus, the null effect of density is an indication that aphasic errors, like normal errors, are strongly constrained by the characteristics of the target itself. This finding is also in complete accord with studies showing that single-phoneme substitution errors are the most frequent type of phonological error in both normal (e.g. Boomer & Laver, 1968; Garnham et al., 1981) and aphasic (e.g. Beland et al., 1990; Blumstein, 1973a) speakers, because words which differ by only one phoneme belong to the same neighbourhood, and will therefore have similar neighbourhood densities.

Despite the lack of a main effect of density, closer inspection suggests that neighbourhood density may not be preserved equally across different types of phonological errors. In Figures 5-iv and 5-vi, it is evident that density shows opposing effects for word and non-word errors, which was statistically supported by a significant interaction between item set (errors *vs* targets) and lexical status of error (words *vs* non-words). Analysis of the interaction revealed that there was no significant density effect for word errors, but that non-word errors were significantly *lower* in density than their targets. Thus, non-word errors seem to represent a less 'word-like' set of items than either their targets, or than word errors and their targets. On the face of it, this may seem like a truism; but lower density values means that the non-word errors have more unusual wordshapes, which in turn implies that they are structurally more complex, or perhaps consist of less frequent, more marked structures. However, the opposite proposal is inherent in hypotheses that neologistic errors represent simplified, less marked, or 'default' productions (e.g. Béland et al., 1993). Rather, it appears that the phonological characteristics constraining word errors are loosened in the production of non-word errors. In a connectionist model, such 'loosening' of constraints might be modeled by increasing the amount of noise in the system, resulting in production which approximates the 'random' end of the spectrum.

Although phonological characteristics are largely preserved in error production, it seems that frequency variables are not. Although strongly correlated with neighbourhood density, the effect of neighbourhood frequency seems to exert a separate effect: among candidate items with similar density values (perhaps from within the same neighbourhood), items with higher neighbourhood frequencies appear to have an advantage. More difficult to explain are the large discrepancies apparent between the raw item frequency values of errors and targets (again see Figures 5-iv and 5-vi), which occur in opposite directions in the two tasks. Log-transformation of the data, however, statistically nullified the difference in the PNT, indicating that outliers were contributing to a great extent to the frequency differential. Despite the transformation, the frequency difference remained in the NR task, such that errors were significantly lower in frequency than their targets, contrary to expectations. As mentioned in the previous chapter, this result may nonetheless represent a statistical artifact of outlying values, but other explanations are also possible.

The simplest explanation is that the results represent real qualitative differences between high- and low-frequency words. Neighbourhood analyses with normal subjects have shown that high-frequency items sometimes show different effects from low-frequency items (Luce & Pisoni, 1998; Vitevitch, 1997). Thus, it may be informative in future analyses to divide the targets into high- and low-frequency categories. Previous studies have taken this step, but the cut-off value between high- and low-frequency items is often arbitrarily determined and varies widely across studies (e.g. see Best, 1996; Stemberger & MacWhinney, 1986). A more motivated approach might be to define natural categories of frequency. For example, separating content and function words would, theoretically, put most of the extremely high-frequency items in a separate category (and would take care of the outliers in the present data). However, the content/function word distinction is not respected by the measure of *phonological* frequency, which is a more relevant measure for phonological neighbourhood studies than lexical frequency, thus making it difficult to isolate function words. An alternative method would be to simply avoid the extremes of the frequency continuum (Jescheniak & Levelt, 1994; Savage et al., 1990); however, frequency information representative of natural language samples is necessarily lost with such an approach.

One other possibility that might explain the apparent substitution of lower frequency words for higher frequency targets relates to the way in which word errors were defined. Whereas the targets represent words which occurred naturally (albeit in error) during the course of picture description, the error words were classified as words on the basis of occurring in the dictionary, which contains many very infrequent words. Thus low-frequency errors such as *braise*, *scape*, and *pyre* were classified as words, and included in the word error set, even though their lexical status may be due to chance. Chance word outcomes cannot be ruled out for high-frequency word errors either, but it seems likely that the lexical status of a rare word which may not even be in the vocabulary of the speaker has little meaning. For the present, the unexpected frequency effect observed between errors and targets must be interpreted with caution.

The possibility remains that the current analysis was simply not sensitive enough to pick up phonological differences between errors and targets. Vitevitch (1997), for example, found no overall differences between errors and targets on frequency, density, or neighbourhood frequency, but did find a frequency difference when comparing the *relative* frequencies of individual errors to their respective targets (see also del Viso, Igoa, & Garcia-Albert, 1991, cited in Vitevitch, 1997). Similar effects have been shown for the relative frequencies of phonemes in errors and targets (Levitt & Healy, 1985). This type of analysis might also be useful to test for differential neighbourhood effects in specific types of errors.

Effects of Lexical Status

The consistent and large differences between word error and non-word error sets are easily explained by reference to the inherent characteristics of the language, and the inter-correlations among the variables. Because high-density targets have more word neighbours by definition, phonological errors are more likely to be words, while low-density targets are more likely to give rise to non-word errors (Best, 1995; Dell & Reich, 1981; Nickels & Howard, 1995). Thus, the lexical status of error effect on neighbourhood density reflects the opportunities provided by the lexicon. Word length is closely associated with density, and tends to be preserved in error production. For example, one-syllable targets, which have higher density neighbourhoods, are likely to give rise to one-syllable errors, which are more likely to be words. Neighbourhood frequency is also dependent on neighbourhood density to some extent, so the lexical status of error effect here is also to be expected.

In sum, although the effects of item frequency on error outcome are questionable, the most important finding about error outcome is that phonological characteristics of the target are, for the most part, preserved in error production, particularly in word errors. The significant facilitative effect of neighbourhood frequency found in both tasks does not put this conclusion into jeopardy, as it is not a phonological characteristic *per se*. The consistent effects of lexical status found in the current study, along with the close correspondence of error and target characteristics, illustrate that the lexical status of the error is largely determined by the nature of the target. In other words, certain types of targets (longer, lower frequency words with sparser, less frequent neighbourhoods) are more likely to give rise to non-word errors. The fact that these are exactly the types of words which result in more errors overall lends support to the hypothesis of Dell and colleagues (1997a) that a greater proportion of non-word errors is a reflection of severity of aphasia. The strong lexical status effects shown in this study illustrate the importance of analyzing word and non-word errors separately.

Support for the Continuity Thesis

The fact that these results parallel results found for normal subjects validates the underlying assumption of the continuity thesis, that aphasic deficits represent more extreme manifestations of the malfunctions which occasionally disrupt normal speech production (Buckingham, 1980; 1999; Dell et al., 1997a; Dell et al., 1997b; Talo, 1980). As in studies of normal speech errors (e.g. Vitevitch, 1997; Vitevitch, ms in prep), the neighbourhood effects shown in the current study can be accounted for within a connectionist interactive spreading activation model, which allows a continuum of performance to be modelled, from normal to random, through global alterations of such parameters as connection weight or rate of activation decay. Numerous studies have illustrated that a variety of aphasic error patterns can be accommodated within such a model (e.g. Croot et al., 1998; Dell et al., 1997b; Harley & MacAndrew, 1992; Laine & Martin, 1996; Martin et al., 1994; Schwartz & Brecher, 2000; Schwartz et al., 1994; Wright & Ahmad, 1997). Preliminary data from the pilot study also supported data from previous studies (Dell et al., 1997a; Schwartz et al., 1994) indicating a relationship between severity of aphasia and the ratio of anticipations to perseverations. In the words of Freud, acknowledged as one of the original proponents of the continuity thesis (Buckingham, 1999; Dell et al., 1997b), "It is tempting to regard paraphasia in the widest sense as a purely functional symptom, a sign of reduced efficiency of the apparatus of speech associations" (Freud, 1901, cited in Buckingham, 1999).

If the facilitative effects of the phonological neighbourhood operate by boosting the activation of the intended target, then aphasic speakers whose word-finding deficits can be characterized as a consequence of weakened connections among lexical nodes (i.e. reduced efficiency) would be expected to show attenuated effects of neighbourhood density and neighbourhood frequency. Deficits due to abnormally rapid rates of activation decay might be expected to have similar consequences on the effects of neighbourhood variables, because the ability of the lexical network to resolve the activation of multiple candidates for selection relies on sufficient processing time (Dell, 1985). On the other hand, it is conceivable that denser, more frequent neighbourhoods might have a detrimental effect on accurate target selection in patients with decay deficits. If the feedback mechanism 'times out' at the decision stage without having selected a word, a greater number of candidates would remain active for high- than low-density items, resulting in a greater likelihood that an error would be produced. Thus, the strength of the neighbourhood effect would be expected to depend in some way on the severity of the lexical access deficit. However, the results of the neighbourhood analyses in the current study do not speak directly to the issue of severity; no severity effects were evident in the individual phonological error rates, and no other types of errors were examined in the main study. Furthermore, the numbers of errors produced by the subjects were not sufficient, in most cases, to assess and compare neighbourhood effects in the individual subject data. In future studies, measuring reaction time of naming response might provide a more sensitive indication of severity than accuracy (Vitevitch, ms in prep).

Nevertheless, the speech production performance of the aphasic subjects as a group represents a significant departure from normality, as seen in the comparison of error rates between aphasic and control groups in the NR task. Despite this quantitative difference, they showed patterns of error production which were qualitatively similar to the control subjects, and neighbourhood effects similar to normal subjects in previous studies, in accordance with the continuity thesis. Target susceptibility results paralleled studies of normal errors (Vitevitch, 1997; Vitevitch, ms in prep), and the error outcomes showed the preservation of phonological constraints, at least for word errors. Even the density effect found for non-word errors, although not initially predicted, is in line with the continuity thesis; a loosening of the constraints governing errors would be expected in errors which reflect a more severe disruption of lexical access. On the contrary, evidence of the operation of default or compensatory mechanisms, reflected in an over-reliance on linguistic rules and constraints, would be counter to the predictions of the continuity thesis; Dell claims that errors become more random as severity of the lexical access deficit increases (Dell et al., 1997b). (Note that this argument makes a distinction between adherence to the constraints determined by the intended target, which reflects normal processing, and *over*-adherence to output constraints determined by regularities and preferred patterns in the language, which reflects pathological, though perhaps compensatory, processing.)

These speculations, however, await further investigation. Only phonologically target-related errors were analyzed in this study; perhaps other types of errors may reveal different effects. For example, unrelated errors might reflect the operation of default mechanisms; if there are *no* target constraints governing an error, compensatory devices such as the generation of high-frequency phoneme sequences by a 'random generator', may reveal an influence on error output. As described in Chapter 2, the influences of different constraints appear to trade off against each other. Analyses of these effects in individual subjects would also be required to determine the relationship between severity and neighbourhood effects. Finally, hypotheses about the continuity thesis which rely on a connectionist paradigm must be tested through computational simulations of the error patterns.

Clinical Implications

From a clinical perspective, adverse effects of neighbourhood density on lexical access might have been expected for aphasics, despite previous findings from normal subjects. In speech-language therapy settings, most aphasic clients show a propensity for phonological confusions, suggesting that phonological relatedness inhibits rather than facilitates access. More generally, clinicians typically observe that a larger set of options in a forced-choice task promotes more errors. However, the results should not be surprising, given the documented success of phonological cueing techniques (e.g. Pease & Goodglass, 1978; Spencer et al., 2000), which are probably used by all clinicians, and have been since aphasics were first treated. The hypothesis of Dell and colleagues (Dell et al., 1997b), that the naming deficits of many aphasic subjects are due to weakened connections within the phonological lexicon, provides an explanation for the mechanism underlying such cueing techniques, which operates by boosting the activation of the target's phonological neighbourhood.

The findings that neighbourhood density and item frequency influence the susceptibility of target words to error adds to our knowledge of the myriad stimulus factors which promote or reduce the accuracy of speech production in aphasia. Thus, phonological neighbourhood variables provide another dimension along which to structure hierarchies of task difficulty in diagnostic and therapeutic intervention. As primarily an investigation of item characteristics, rather than subject characteristics, the present study serves as a starting point for more specific investigations. More research is required to test hypotheses about the impact of neighbourhood variables on specific types of errors, and their role in different clinical sub-types of aphasia. For example, do phonological neighbourhoods also play a role in the production of non-phonologically related errors? Can analogous semantic neighbourhoods be modelled and tested for similar effects? Are phonological neighbourhood effects evident in aphasic subjects with phonological planning deficits as well as those with phonological activation deficits?

More generally, the role of interactive activation models and the concept of the continuity thesis may guide the way in which speech-language pathologists view aphasic deficits, and thus the types of intervention which are used to address these deficits. Findings that seemingly specific deficits can result from global processing impairments emphasize the need to minimize processing load during language tasks, for example, by reducing distractions and controlling for the patient's level of fatigue during therapy. Distinguishing between connection-weight and decay impairments can indicate whether patients will benefit from a faster or slower rate of stimulus presentation (see Brookshire, 1992). Efforts to model generalization of behaviours and aphasic recovery have also successfully used connectionist approaches (e.g. Harley, 1996). Connectionist approaches stress the probabilistic, gradational nature of language breakdown, and are thus fully compatible with clinical observations of inconsistencies in aphasic performance, overlapping symptoms among aphasic sub-types, and differential recovery patterns, which have been so problematic for modular explanations of aphasic behaviour (Schwartz, 1984).

Future Research

Neighbourhood Definitions

In addition to the follow-up analyses suggested in the preceding discussion, a major issue to be addressed in order to corroborate the null findings of neighbourhood variables on error outcome concerns the nature of the phonological neighbourhood. The specific characteristics of the neighbourhood, and how they exert their influence, remain speculative. In modeling the lexicon, it is necessary to make some assumptions, which at this point are necessarily somewhat arbitrary, about the contents of the phonological lexicon. In the Neighborhood Activation Model, the phonological neighbourhood is defined as those lexical items differing from the stimulus by one phoneme; similarly, in the model of Dell et al. (1997b), which included only CVC words, "[e]ach word's formal neighbors matched on exactly two phonemes" but they note that "In real neighbourhoods, semantic and phonological similarity is graded" (p. 832).

How one defines a 'real neighbourhood' should begin with an analysis of the relatedness effects observed between errors and their targets. The prevalence of single-phoneme errors in normal and aphasic corpora (e.g. Blumstein, 1973a; Garnham et al., 1981) provides strong support for the current neighbourhood definition. However, many phonological errors diverge from their targets by more than one phoneme, yet are still classified as 'target-related' (see Kohn & Smith, 1994a for one definition). In the current study, only one phoneme overlap was minimally required for an error to qualify as phonologically related to its target.

Relatedness may also go beyond the degree of phonological overlap. For example, studies have shown that phonologically related errors may share stress pattern but few phonemes with their targets (Fay & Cutler, 1977; Kohn & Smith, 1994a). Furthermore, spoken word recognition studies have shown relatedness effects between primes and targets which share phonetic features but have no phonemes in common (Goldinger et al., 1989; Goldinger et al., 1992). More weight may need to be assigned to syllabic frequency effects (e.g. Dell, 1986; Levelt & Wheeldon, 1994; Sussman, 1984). In addition, overlap in some word positions, onsets in particular, may be more important than overlap in other positions (e.g. Marslen-Wilson & Zwitserlood, 1989; Shattuck-Hufnagel, 1987). All these factors may contribute differentially to error outcome, given the apparent trading off of constraints observed in error production.

In examining relatedness effects between errors and targets, one must be especially vigilant about the reliability of the transcription. The current study required a relatively gross level of analysis; a decision was required only as to whether or not the error and target were phonologically related. However, discerning the nature of the relationship requires much more specific decisions. Thus perceptual ratings should be supplemented with objective (i.e. acoustic or physiological) measurements (Kent, 1996; Mowrey & MacKay, 1990; Shinn & Blumstein, 1983). In the words of Buckingham and Yule (1987), "good phonetics can be used to avoid bad phonology or, more specifically, false evaluations" (p. 115).

Error Elicitation Experiments

Confirmation of the present results will be important, by using error elicitation paradigms similar to those which have been used for normal subjects. In particular, the replication of neighbourhood studies using the same tasks, such as the elicitation of tip-of-the-tongue states (Harley & Bown, 1998), the induction of spoonerisms (Vitevitch, ms in prep, Exp. 1 & 2), or the analysis of naming latencies (Vitevitch, ms in prep, Exp. 3) would provide a more stringent test of the continuity thesis, and the ability of connectionist interactive models to account for aphasic deficits. In these studies, stimuli are divided into groups based on high and low values of the neighbourhood variables. This method allows greater control over the variables of interest, avoiding some of the potential problems encountered in the current study, such as an insufficient range of density values, and skewness in the frequency distributions. In combination with more naturalistic tasks such as those used here, such an experiment would provide corroborative evidence of neighbourhood effects.

Another paradigm adopted from studies with normal subjects is the picture-word interference task (e.g. Schriefers et al., 1990). In this task, pictures are presented for naming, and within each trial, a distractor word is also presented, either auditorily or in written form, which may be semantically related, phonologically related or unrelated to the picture. As in Stroop tasks, when the distractor is semantically related to but does not match the target, it has been shown to interfere with the naming response (thus slowing reaction time); in contrast, phonologically related 'distractors' have been shown to facilitate reaction time (e.g. Schriefers et al., 1990; see Glaser, 1992 for a review). The use of distractors allows the assessment of different types and degrees of relatedness on the response and, thus, may provide a means of exploring the nature of the phonological lexicon. In addition, as mentioned earlier, reaction times may provide a more sensitive method of measurement than accuracy (Vitevitch, ms in prep). The pilot study and the main study presented here have provided strong indications of the role of the phonological neighbourhood in lexical access in aphasia. Such elicitation experiments would complement these findings with converging evidence from more structured tasks.

Conclusions

Normal speech errors have long been a source of curiosity and entertainment, from the legendary Reverend Spooner and the fictitious Mrs. Malaprop to the every-day slips of the tongue committed by us all. Likewise, the occurrence of aphasic speech errors has often injected a welcome note of comic relief, not least of all for the patient, into many a therapy session. However, paraphasic errors are just as frequently a source of frustration and embarrassment for the patient. As one of the most obvious and debilitating manifestations of aphasia, the analysis of paraphasias has been a focus of investigations into the underlying deficits of aphasia for over a century. Such investigations, along with studies of normal errors, have revealed a certain "systematicity and regularity" (Blumstein, 1991, p. 158) of error production, due to a number of linguistic constraints which induce certain linguistic structures to 'slip' in certain contexts, or which influence the nature of the slip itself. As Blumstein notes, "the performance of the patients is variable, but it is NOT random, as it follows specific phonological principles" (p. 158).

Nevertheless, error studies have frequently given rise to apparently contradictory evidence. This seems, in part, to be due to the complex interactions and trade-offs among myriad constraints operative in different contexts. On-going speech-error research continues to clarify the processes involved in lexical retrieval (and mis-retrieval), by combining experimental approaches and theoretical perspectives from the fields of psychology and linguistics. According to Dell (1988), whereas the linguistic tradition (exemplified by Blumstein above) emphasizes linguistic rules and constraints, the psychological tradition emphasizes the probabilistic interaction of multiple influences. Current connectionist computational model techniques (exemplified by Dell's (1986) interactive activation model of speech production) has allowed the testing of multiple influences on lexical access through computer simulations.

The Neighbourhood Activation Model (NAM, Luce & Pisoni, 1998) was proposed to explain the influence of three such probabilistic factors on lexical access in speech recognition: frequency of occurrence, neighbourhood density, and neighbourhood frequency. Whereas these were shown to have competitive effects in word recognition, they have recently been found to facilitate accurate speech production in normal subjects (Vitevitch, ms in prep), a finding which was replicated in the current study. Modifications to the NAM are necessary for it to be able account for the data from production studies. Neighbourhood variables have not been shown to influence error outcomes, however, in any systematic way, either in previous research with normal subjects (Vitevitch, 1997; Vitevitch, ms in prep), or in the present study. Rather, the finding that errors apparently *preserve* many of the characteristics of their targets supports the continuity hypothesized between normal and aphasic error patterns (Dell et al., 1997a). However, more research is required to validate the current working definition of the phonological neighbourhood.

References

Alajouanine, T. (1956). Verbal realisation in aphasia. *Brain, 79*, 1-28.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(5), 802-814.

Ardila, A., & Rosselli, M. (1993). Language deviations in aphasia: A frequency analysis. *Brain and Language, 44*, 165-180.

Baars, B. J. (1980). The competing plans hypothesis: An heuristic viewpoint on the causes of errors in speech. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler* (pp. 39-49). The Hague: Mouton.

Baars, B. J. (1992a). Ch. 6: A dozen competing-plans techniques for inducing predictable slips in speech and action. In B. J. Baars (Ed.), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (pp. 129-150). New York, NY: Plenum Press.

Baars, B. J. (Ed.). (1992b). *Experimental Slips and Human Error: Exploring the Architecture of Volition.* New York, NY: Plenum Press.

Baars, B. J., Motley, M. T., & MacKay, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14*, 382-391.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two-stage model of lexical retrieval: Evidence from a case of anomia with selective preservation of grammatical gender. *Cognition, 57*, 193-216.

Barton, M., Maruszewski, M., & Urrea, D. (1969). Variation of stimulus context and its effect on word-finding ability in aphasics. *Cortex, 5*, 351-365.

Basso, A., Razzano, C., Faglioni, P., & Zanobio, M. E. (1990). Confrontation naming, picture description and action naming in aphasic patients. *Aphasiology, 4*(2), 185-195.

Bastiaanse, R., Gilbers, D., & van der Linde, K. (1994). Sonority substitutions in Broca's and conduction aphasia. *Journal of Neurolinguistics, 8*, 247-255.

Baum, S. R. (1997). Phonological, semantic, and mediated priming in aphasia. *Brain and Language, 60*, 347-359.

Baum, S. R., Blumstein, S. E., Naeser, M. A., & Palumbo, C. L. (1990). Temporal dimensions of consonant and vowel production: An acoustic and CT scan analysis of aphasic speech. *Brain and Language, 39*, 33-56.

Baum, S. R., & Slatkovsky, K. (1993). Phonemic false evaluation? Preliminary data from a conduction aphasic patient. *Clinical Linguistics & Phonetics, 7*(3), 207-218.

Bawden, H. H. (1900). A study of lapses. *Psychological Monographs, 3*, 1-121.

Beattie, G. W., & Butterworth, B. L. (1979). Contextual probability and word frequency as determinants of pauses and errors in spontaneous speech. *Language and Speech, 22*(3), 201-211.

Béland, R. (1998). Feature geometry and phonemic paraphasias. In P. Bhatt & R. Davis (Eds.), *The Linguistic Brain* (pp. 127-142). Toronto, ON: Canadian Scholars' Press.

Béland, R., Caplan, D., & Nespoulos, J.-L. (1990). The role of abstract phonological representations in word production: Evidence from phonemic paraphasias. *Journal of Neurolinguistics, 5*, 125-164.

Béland, R., Paradis, C., & Bois, M. (1993). Constraints and repairs in aphasic speech: A group study. *Canadian Journal of Linguistics/Revue canadienne de linguistique, 38*(2), 279-302.

Best, W. M. (1995). A reverse length effect in dysphasic naming: When elephant is easier than ant. *Cortex, 31*, 637-652.

Best, W. M. (1996). When racquets are baskets but basket are biscuits, where do words come from? A single case study of formal paraphasic errors in aphasia. *Cognitive Neuropsychology, 13*(3), 443-480.

Bierwisch, M. (1981). Linguistics and language error. *Linguistics, 19*, 583-626.

Blanken, G. (1990). Formal paraphasia: A single case study. *Brain and Language, 38*(4), 534-554.

Blanken, G., Dittman, J., Haas, J.-C., & Wallesch, C. W. (1988). Producing speech automatisms (recurring utterances): Looking for what is left. *Aphasiology, 2*(6), 545-556.

Blumstein, S. E. (1973a). *A Phonological Investigation of Aphasic Speech*. The Hague: Mouton.

Blumstein, S. E. (1973b). Some phonological implications of aphasic speech. In H. Goodglass & S. Blumstein (Eds.), *Psycholinguistics and Aphasia* (pp. 123-137). Baltimore, MD: The Johns Hopkins University Press.

Blumstein, S. E. (1990). Phonological deficits in aphasia: Theoretical perspectives. In A. Caramazza (Ed.), *Cognitive Neuropsychology and Neurolinguistics: Advances in Models of Cognitive Function and Impairment* . Hillsdale, NJ: Lawrence Erlbaum.

Blumstein, S. E. (1991). Phonological aspects of aphasia, *Acquired Aphasia* (second ed.). New York, NY: Academic Press.

Blumstein, S. E., Cooper, W. D., Goodglass, H., Statlender, S., & Gottlieb, J. (1980). Production deficits in aphasia: A voice-onset time analysis. *Brain and Language, 9*, 153-170.

Blumstein, S. E., Cooper, W. D., Zurif, E. B., & Caramazza, A. (1977). The perception and production of voice-onset-time in aphasia. *Neuropsychologia, 15*, 371-383.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language and Communication* (pp. 181-216). San Diego, CA: Academic Press.

Bock, K., & Levelt, W. J. M. (1994). Language production. In M. A. Gernsbacher (Ed.), *The Handbook of Linguistics* (pp. 945-984). San Diego, CA: Academic Press.

Boomer, D. S., & Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication, 3*, 2-12.

Boyczuk, J. P., & Baum, S. R. (1999). The influence of neighborhood density on phonetic categorization. *Brain and Language, 67*(1), 46-70.

Brookshire, R. H. (1992). *An Introduction to Neurogenic Communication Disorders.* (4th ed.). St. Louis, MO: Mosby Year Book.

Browman, C. P. (1980). Ch. 15: Perceptual processing: Evidence from slips of the ear. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 213-230). New York, NY: Academic Press.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109*(2), 204-223.

Brown, L. (Ed.). (1993). *The New Shorter Oxford English Dictionary*. Oxford, England: Clarendon Press.

Brown, R., & McNeill, D. (1966). The tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325-337.

Buckingham, H., & Kertesz, A. (1974). A linguistic analysis of fluent aphasia. *Brain and Language, 1*, 43-62.

Buckingham, H. W. (1977). The conduction theory and neologistic jargon. *Language and Speech, 20*(2), 174-184.

Buckingham, H. W. (1980). On correlating aphasic errors with slips of the tongue. *Applied Psycholinguistics, 1*, 199-220.

Buckingham, H. W. (1981). Where do neologisms come from? In J. W. Brown (Ed.), *Jargonaphasia* (pp. 39-62). New York, NY: Academic Press.

Buckingham, H. W. (1986). The scan-copier mechanism and the positional level of language production: Evidence from phonemic paraphasia. *Cognitive Science, 10*, 195-217.

Buckingham, H. W. (1987). Phonemic paraphasias and psycholinguistic production models for neologistic jargon. *Aphasiology, 1*, 381-400.

Buckingham, H. W. (1990). Abstruse neologisms, retrieval deficits and the random generator. *Journal of Neurolinguistics, 5*, 215-235.

Buckingham, H. W. (1992). Phonological production deficits in conduction aphasia. In S. E. Kohn (Ed.), *Conduction Aphasia* (pp. 77-116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Buckingham, H. W. (1999). Freud's continuity thesis. *Brain and Language, 69*(1), 76-92.

Buckingham, H. W., Avakian-Whitaker, H., & Whitaker, H. A. (1978). Alliteration and assonance in neologistic jargon aphasia. *Cortex, 14*, 365-380.

Buckingham, H. W., & Yule, G. (1987). Phonemic false evaluation: Theoretical and clinical aspects. *Clinical Linguistics & Phonetics, 1*(2), 113-125.

Burns, M. S., & Canter, G. J. (1977). Phonemic behavior of aphasic patients with posterior cerebral lesions. *Brain and Language, 4*, 492-507.

Burton, M. W. (1989). *Associative, mediated, and rhyme priming effects: A study of lexical processing.* , Brown University, Providence, RI.

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 567-575.

Butterworth, B. (1979). Hesitation and the production of verbal paraphasias and neologisms in jargon aphasia. *Brain and Language, 8*, 133-161.

Butterworth, B. (1980). Some constraints on models of language production. In B. Butterworth (Ed.), *Language Production* (Vol. 1: Speech and Talk). London, Eng.: Academic Press.

Butterworth, B. (1981). Speech errors: Old data in search of new theories. *Linguistics, 19*, 627-662.

Butterworth, B. (1992). Disorders of phonological encoding. *Cognition, 42*, 261-286.

Canter, G., Trost, J., & Burns, M. (1985). Contrasting speech patterns in apraxia of speech and phonemic paraphasia. *Brain and Language, 24*, 204-222.

Caplan, D. (1987). *Neurolinguistics and Linguistic Aphasiology: An Introduction.* (Cambridge Studies in Speech Science and Communication). Cambridge, NY: Cambridge University Press.

Cappa, S., Cavallotti, G., & Vignolo, L. A. (1981). Phonemic and lexical errors in fluent aphasia: Correlation with lesion site. *Neuropsychologia, 19*(2), 171-177.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177-208.

Caramazza, A., & Miozzo, M. (1997). The relation between syntactic and phonological knowledge in lexical access: Evidence from 'tip-of-the-tongue' phenomenon. *Cognition, 64*, 309-343.

Christman, S. S. (1994). Target-related neologism formation in jargonaphasia. *Brain and Language, 46*, 109-128.

Code, C. (1982). Neurolinguistic analysis of recurrent utterances in aphasia. *Cortex, 18*, 141-152.

Code, C., & Ball, M. J. (1994). Syllabification in aphasic recurring utterances: Contributions of sonority theory. *Journal of Neurolinguistics, 8*, 257-265.

Cohen, A. (1980). Correcting of speech errors in a shadowing task. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 158-164). London, Eng.: Academic Press.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64*, 44-56.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Domic (Ed.), *Attention and Performance VI* (pp. 535-555). Hillsdale, NJ: Erlbaum.

Columbo, L. (1986). Activation and inhibition with orthographically similar words. *Journal of Experimental Psychology: Human Perception and Performance, 12*(2), 226-234.

Croot, K., Patterson, K., & Hodges, J. R. (1998). Single word production in nonfluent progressive aphasia. *Brain and Language, 61*(2), 226-273.

Cutler, A. (1980). Ch. 4: Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 67-80). New York, NY: Academic Press.

Cutler, A. (1981). The reliability of speech error data. *Linguistics, 19*, 561-582.

Cutler, A. (1995). Spoken word recognition and production. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language and Communication* (pp. 97-136). San Diego, CA: Academic Press.

del Viso, S., Igoa, J. M., & Garcia-Albert, J. E. (1991). On the autonomy of phonological encoding: Evidence from slips of the tongue in Spanish. *Journal of Psycholinguistic Research, 20*, 161-185.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 222-233.

Dell, G. S. (1985). Positive feedback in hierarchical connectionist models: Applications to language production. *Cognitive Science, 9*, 3-23.

Dell, G. S. (1986). A spreading activation theory of retrieval in speech production. *Psychological Review, 93*, 283-321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27*, 124-142.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5*(4), 313-349.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997a). Language production and serial order: A functional analysis and a model. *Psychological Review, 104*(1), 123-147.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17*, 149-195.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98*, 604-614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*, 287-314.

Dell, G. S., & Reich, P. A. (1981). Stages in sentence production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20*, 611-629.

Dell, G. S., & Repka, R. J. (1992). Ch. 10: Errors in inner speech. In B. J. Baars (Ed.), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (pp. 237-262). New York, NY: Plenum Press.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997b). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801-838.

Dressler, W. U. (1979). Experimentally induced phonological paraphasias. *Brain and Language, 8*(1), 19-24.

Ellis, A. (1980). On the Freudian theory of speech errors. In V. Fromkin (Ed.), *Errors in Linguistic Performance* (pp. 123-132). New York, NY: Academic Press.

Ellis, A. (1985). The production of spoken words: A cognitive neuropsychological perspective. In A. Ellis (Ed.), *Progress in the Psychology of Language* (Vol. 2, pp. 107-145). London: Lawrence Erlbaum Associates.

Ellis, A. W., Miller, D., & Sin, G. (1983). Wernicke's aphasia and normal language processing: A case study in cognitive neuropsychology. *Cognition, 15*, 111-144.

Favreau, Y., Nespoulos, J.-L., & Lecours, A. R. (1990). Syllable structure and lexical frequency effects in the phonemic errors of four aphasics. *Journal of Neurolinguistics, 5*, 165-187.

Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry, 8*, 505-520.

Ferber, R. (1991). Slip of the tongue or slip of the ear? One the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research, 20*, 105-122.

Forster, K. I., & Taft, M. (1994). Bodies, antibodies, and neighborhood-density effects in masked form priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 844-863.

Forster, K. L., & Davis, C. (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.

Fowler, C. A. (1995). Speech production. In J. L. Miller & P. D. Eimas (Eds.), *Speech, Language and Communication* (pp. 29-61). San Diego, CA: Academic Press.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 526-540.

Francis, W. N., & Kučera, H. (1982). *Frequency Analysis of English Usage*. Boston, MA: Houghton Mifflin Company.

Freud, S. (1901). *Zur Psychopathologie des Alltaglebens*. (10 ed.). Leipzig: Internationaler Psychoanalytischer Verlag.

Freud, S. (1965). *Psychopathology of everyday life* (Tyson, A., Trans.). New York, NY: Norton.

Friedman, & Kohn, S. E. (1990). Impaired activation of the phonological lexicon: Effects upon oral reading. *Brain and Language, 38*, 278-297.

Fromkin, V. (1993). Speech production. In J. Berko Gleason & N. Bernstein Ratner (Eds.), *Psycholinguistics* (pp. 271-300). Fort Worth: Harcourt-Brace-Jovanovitch.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47*, 27-52.

Fromkin, V. A. (Ed.). (1980). *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand*. New York, NY: Academic Press.

Gagnon, D. A., & Schwartz, M. F. (1996). *The origins of neologisms in picture naming by fluent aphasics*. Paper presented at the TENNET VII, Montréal, QC.

Gagnon, D. A., & Schwartz, M. F. (1997). *Serial position effects in aphasics' neologisms*. Paper presented at the Academy of Aphasia, Philadelphia, PA.

Gagnon, D. A., Schwartz, M. F., Martin, N., Dell, G. S., & Saffran, E. M. (1997). The origins of formal paraphasias in aphasics' picture naming. *Brain and Language, 59*(3), 450-472.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6*, 110-125.

Garnes, S., & Bond, Z. S. (1980). Ch. 16: A slip of the ear: A snip of the ear? A slip of the year? In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 231-239). New York, NY: Academic Press.

Garnham, A., Shillcock, R. C., Brown, G. D. A., Mill, A. I. D., & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics, 19*, 805-817.

Garnsey, S. M., & Dell, G. S. (1984). Some neurolinguistic implications of prearticulatory editing in production. *Brain and Language, 23*, 67-73.

Garrett, M. (1992). Disorders of lexical selection. *Cognition, 42*, 143-180.

Garrett, M. F. (1975). The analysis of sentence production. In G. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 9). New York, NY: Academic Press.

Garrett, M. F. (1976). Syntactic processes in sentence production. In R. J. Wales & E. C. T. Walker (Eds.), *New Approaches to Language Mechanisms* (pp. 231-255). Amsterdam: North Holland.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language Production* (pp. 177-220). London, Eng.: Academic Press.

Garrett, M. F. (1984). The organization of processing structure for language production. In D. Caplan, A. R. Lecours, & A. Smith (Eds.), *Biological Perspectives on Language* (pp. 172-193). Cambridge, MA: MIT Press.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (Vol. III, Language: Psychological and Biological Aspects). Cambridge: Cambridge University Press.

Glaser, W. R. (1992). Picture naming. *Cognition, 42*, 61-105.

Goldberg, E., & Obler, L. K. (1997). *Affixational morphology in jargonaphasia.* Paper presented at the Academy of Aphasia, Philadelphia, PA.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*, 501-518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1211-1238.

Goodglass, H., Hyde, M. R., & Blumstein, S. (1969). Frequency, picturability and availability of nouns in aphasia. *Cortex 2*, 104-119.

Goodglass, H., & Kaplan, E. (1983). *The Assessment of Aphasia and Related Disorders.* Philadelphia, PA: Lea & Febiger.

Goodglass, H., Kaplan, E., Weintraub, S., & Ackerman, N. (1976). The tip-of-the-tongue phenomenon in aphasia. *Cortex, 12*, 145-153.

Goodglass, H., Quadfasel, F. A., & Timberlake, W. H. (1964). Phrase length and type and severity of aphasia. *Cortex, 1*, 133-153.

Goodglass, H., Wingfield, A., Hyde, M. R., Gleason, J. B., Bowles, N. L., & Gallagher, R. E. (1997). The importance of word-initial phonology: Error patterns in prolonged naming efforts by aphasic patients. *Journal of the International Neuropsychological Society, 3*, 128-138.

Gordon, J. K. (1998). The fluency dimension in aphasia. *Aphasiology, 12*(7/8), 673-688.

Gordon, J. K., & Baum, S. R. (1994). Rhyme priming in aphasia: The role of phonology in lexical access. *Brain and Language, 47*, 661-683.

Grainger, J. (1990). Word frequency and neighborhood frequency effects in lexical decision and naming. *Journal of Memory and Language, 29*, 228-244.

Green, E. (1969). Phonological and grammatical aspects of jargon in an aphasic patient: A case study. *Language and Speech, 12*(2), 103-118.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267-283.

Guralnik, D. B. (Ed.). (1986). *Webster's New World Dictionary* (2nd College ed.). New York, NY: Simon and Schuster.

Harley, T. (1996). Connectionist modeling of the recovery of language functions following brain damage. *Brain and Language, 52*, 7-24.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*, 191-219.

Harley, T. A. (1990). Environmental contamination of normal speech. *Applied Psycholinguistics, 11*, 45-72.

Harley, T. A. (1993a). Connectionist approaches to language disorders. *Aphasiology, 7*, 221-249.

Harley, T. A. (1993b). Phonological activation of semantic competitors during lexical access in speech production. *Language and Cognitive Processes, 8*(3), 291-309.

Harley, T. A., & Bown, H. E. (1998). What causes a tip-of-the-tongue state? Evidence for lexical neighbourhood effects in speech production. *British Journal of Psychology, 89*, 151-174.

Harley, T. A., & MacAndrew, S. B. G. (1992). Modeling paraphasias in normal and aphasic speech. *Proceedings of the 14th Annual Conference of the Cognitive Science Society, 14*, 378-383.

Hillinger, M. L. (1980). Priming effects with phonemically similar words: The encoding-bias hypothesis reconsidered. *Memory & Cognition, 8*(2), 115-123.

Hockett, C. F. (1967). Where the tongue slips, there slip I, *To Honor Roman Jakobson* . The Hague: Mouton.

Hofmann, E. (1980). Speech control and paraphasia in fluent and nonfluent aphasics. In H. W. Dechert & M. Raupach (Eds.), *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler* (pp. 121-127). The Hague: Mouton.

Hotopf, W. H. N. (1980). Ch. 7: Semantic similarity as a factor in whole-word slips of the tongue. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 97-110). New York, NY: Academic Press.

Howard, D., & Orchard-Lisle, V. M. (1984). On the origin of semantic errors in naming: Evidence from the case of a global dysphasic. *Cognitive Neuropsychology, 1*, 163-190.

Howard, D., Patterson, K., Franklin, S., Orchard-Lisle, V., & Morton, J. (1985). The facilitation of picture naming in aphasia. *Cognitive Neuropsychology, 2*, 49-80.

Howes, D. (1964). Application of the word-frequency concept to aphasia. In A. V. S. DeReuck & M. O'Connor (Eds.), *Disorders of Language* (pp. 47-78). London, Eng.: J.A. Churchill.

Howes, D. H. (1954). On the interpretation of word frequency as a variable affecting speech recognition. *Journal of Experimental Psychology, 48*, 106-112.

Huttenlocher, J., & Kubicek, L. F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*(3), 486-496.

Jakobson, R. (1956). Two aspects of language and two types of aphasic disturbances. In R. Jakobson & M. Halle (Eds.), *Fundamentals of Language* . The Hague: Mouton.

Jakobson, R. (1964). Towards a linguistic typology of aphasia impairments. In A. V. S. DeReuck & M. O'Connor (Eds.), *Disorders of Language* . London, Eng.: Churchill.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 824-843.

Joanette, Y., Keller, E., & Lecours, A. R. (1980). Sequences of phonemic approximations in aphasia. *Brain and Language, 11*, 30-44.

Jones, G. V. (1989). Back to Woodworth: Role of interlopers in the tip-of-the-tongue phenomenon. *Memory & Cognition, 17*, 69-76.

Kay, J., & Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain, 110*, 613-629.

Kempen, G., & Huijbers, P. (1983). The lexicalization process in sentence production and naming: Indirect election of words. *Cognition, 14*, 185-209.

Kent, R. D. (1996). Hearing and believing: Some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology, 5*, 7-23.

Kerschensteiner, M., Poeck, K., & Brunner, E. (1972). The fluency-nonfluency dimension in the classification of aphasic speech. *Cortex, 8*, 233-247.

Kertesz, A., & Benson, D. (1970). Neologistic jargon aphasia: A clinicopathological study. *Cortex, 6*, 362-386.

Kohn, S. E. (1984). The nature of the phonological disorder in conduction aphasia. *Brain and Language, 23*, 97-115.

Kohn, S. E. (1989). The nature of the phonemic string deficit in conduction aphasia. *Aphasiology, 3*, 209-239.

Kohn, S. E., & Goodglass, H. (1985). Picture-naming in aphasia. *Brain and Language, 24*, 266-283.

Kohn, S. E., Melvold, J., & Shipper, V. (1998). The preservation of sonority in the context of impaired lexical-phonological output. *Aphasiology, 12*(4-5), 375-398.

Kohn, S. E., Melvold, J., & Smith, K. L. (1995). Consonant harmony as a compensatory mechanism in fluent aphasic speech. *Cortex, 31*, 747-756.

Kohn, S. E., & Smith, K. L. (1990). Between-word speech errors in conduction aphasia. *Cognitive Neuropsychology, 7*, 133-156.

Kohn, S. E., & Smith, K. L. (1993). Lexical-phonological processing of functors: Evidence from fluent aphasia. *Cortex, 29*, 53-64.

Kohn, S. E., & Smith, K. L. (1994a). Distinctions between two phonological output deficits. *Applied Psycholinguistics, 15*, 75-95.

Kohn, S. E., & Smith, K. L. (1995). Serial effects of phonemic planning during word production. *Aphasiology, 9*(3), 209-222.

Kohn, S. E., Smith, K. L., & Alexander, M. P. (1992). A longitudinal comparison of the oral reading and repetition of nouns in acute fluent aphasics. *Aphasiology, 6*, 397-401.

Kohn, S. E., Smith, K. L., & Alexander, M. P. (1996). Differential recovery from impairment to the phonological lexicon. *Brain and Language, 52*, 129-149.

Kučera, H., & Francis, W. N. (1967). *Computational Analysis of Present-Day American English.* Providence, RI: Brown University Press.

Kupin, J. J. (1982). *Tongue Twisters as a Source of Information.* Unpublished doctoral dissertation, Indiana University, Bloomington.

Laine, M., Kujala, P., Niemi, J., & Uusipaikka, E. (1992). On the nature of naming difficulties in aphasia. *Cortex, 28*, 537-554.

Laine, M., & Martin, N. (1996). Lexical retrieval deficit in picture naming: Implications for word production models. *Brain and Language, 53*(3), 283-314.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119-131.

Laubstein, A. S. (1987). Syllable structure: The speech error evidence. *Canadian Journal of Linguistics, 32*(4), 339-363.

Laubstein, A. S. (1999). Lemmas and lexemes: The evidence from blends. *Brain and Language, 68*, 135-143.

Lavorel, P. M. (1982). Production strategies: A systems approach to Wernicke's aphasia, *Neural Models of Language Processes* (pp. 135-164).

Lecours, A. R. (1982). Simulation of speech production without a computer. In M. A. Arbib, D. Caplan, & J. C. Marshall (Eds.), *Neural Models of Language Processes* (pp. 345-370). New York, NY: Academic Press.

Lecours, A. R. (1990). Introduction: Atheus, Shenkius *et alii. Journal of Neurolinguistics (special issue on phonology in aphasia), 5*(2/3), 113-124.

Lecours, A. R., & Caplan, D. (1975). Book Review: A Phonological Investigation of Aphasic Speech (S.E. Blumstein). *Brain and Language, 2*, 237-254.

Lecours, A. R., & Rouillon, F. (1976). Neurolinguistic analysis of jargonaphasia and jargonagraphia. In H. Whitaker & H. A. Whitaker (Eds.), *Studies in Neurolinguistics* (Vol. 2, pp. 95-144). New York, NY: Academic Press.

Lecours, R., & Lhermitte, F. (1969). Phonemic paraphasias: Linguistic structures and tentative hypotheses. *Cortex, 5*, 193-228.

Levelt, P. (1989). *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Levelt, W. J., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition, 50*, 239-269.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42*, 1-22.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991a). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98*, 122-142.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991b). Normal and deviant lexical processing: Reply to Dell and O'Seaghdha (1991). *Psychological Review, 98*(4), 615-618.

Levitt, A. G., & Healy, A. F. (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language, 24*, 717-733.

Liederman, J., Kohn, S., Wolf, N., & Goodglass, H. (1983). Lexical creativity during instances of word-finding difficulty: Broca's vs. Wernicke's aphasia. *Brain and Language, 20*, 21-32.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception and Psychophysics, 39*(3), 155-158.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear & Hearing*, 1-36.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words, In G. T. M. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 122-147). Cambridge, MA: MIT Press.

MacKay, D. G. (1970a). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8*, 323-350.

MacKay, D. G. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3*, 210-227.

MacKay, D. G. (1980). Ch. 22: Speech errors: Retrospect and prospect. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 319-332). New York, NY: Academic Press.

MacKay, D. G. (1987). *The Organization of Perception and Action: A Theory for Language and other Cognitive Skills*. New York, NY: Springer-Verlag.

MacMahon, M. K. C. (1971). Modern linguistics and aphasia. *British Journal of Disorders of Communication, 7*(1), 54-63.

Marslen-Wilson, W., Moss, H. E., & Van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1376-1392.

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15*(3), 576-585.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29-63.

Martin, A. D., Wasserman, N. H., Gilden, L., Gertman, L., & West, J. A. (1975). A process model of repetition in aphasia: An investigation of phonological and morphological interactions in aphasic error performance. *Brain and Language, 2*, 434-450.

Martin, N., Dell, G. S., Saffran, E. M., & Schwartz, M. F. (1994). Origins of paraphasias in deep dysphasia: Testing the consequences of a decay impairment to an interactive spreading activation model of lexical retrieval. *Brain and Language, 47*(4), 609-660.

Martin, N., Gagnon, D. A., Schwartz, M. F., Dell, G. S., & Saffran, E. M. (1996). Phonological facilitation of semantic errors in normal and aphasic speakers. *Language and Cognitive Processes, 11*(3), 257-282.

Martin, N., Roach, A., Brecher, A., & Lowery, J. (1998). Lexical retrieval mechanisms underlying whole-word perseveration errors in anomic aphasia. *Aphasiology, 12*(4/5), 319-333.

Martin, N., & Saffran, E. M. (1992). A computational account of deep dysphasia: Evidence from a single case study. *Brain and Language, 43*(2), 240-274.

Martin, N., Weisberg, R. W., & Saffran, E. M. (1989). Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. *Journal of Memory and Language, 28*, 462-485.

Mattson, M. E., & Baars, B. J. (1992). Ch. 11: Error-minimizing mechanisms: Boosting or editing? In B. J. Baars (Ed.), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (pp. 263-287). New York, NY: Plenum Press.

Meringer, R., & Mayer, K. (1895). *Versprechen und Verlesen, eine psychologisch linguistische Studie*. Vienna.

Meyer, A. S. (1992). Investigation of phonological encoding through speech error analyses: Achievements, limitations, and alternatives. *Cognition, 42*, 181-211.

Meyer, A. S., & Bock, K. (1992). The tip-of-the-tongue phenomenon: Blocking or partial activation? *Memory & Cognition, 20*(6), 715-726.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition, 2*(2), 309-321.

Miceli, G., Gainotti, G., & Caltagirone, C. (1980). Some aspects of phonological impairment in aphasia. *Brain and Language, 11*, 159-169.

Milberg, W., Blumstein, S., & Dworetsky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society, 26*(4), 305-308.

Miller, D., & Ellis, A. W. (1987). Speech and writing errors in "neologistic jargon aphasia": A lexical activation hypothesis. In M. Coltheart, G. Sartori, & R. Job (Eds.), *The Cognitive Neuropsychology of Language* . Hillsdale, NJ: Erlbaum.

Mitchum, C. C., Ritgert, B. A., Sandson, J., & Berndt, R. S. (1990). The use of response analysis in confrontation naming. *Aphasiology, 4*(3), 261-280.

Moerman, C., Corluy, R., & Meersman, W. (1983). Exploring the aphasiac's naming disturbances: A new approach using the neighbourhood limited classification method. *Cortex, 19*, 529-543.

Monoi, H., Fukusako, Y., Itoh, M., & Sasanuma, S. (1983). Speech sound errors in patients with conduction and Broca's aphasia. *Brain and Language, 20*, 175-194.

Motley, M. T., & Baars, B. J. (1975). Encoding sensitivities to phonological markedness and transitional probability: Evidence from spoonerisms. *Human Communication Research, 1*(4), 353-361.

Motley, M. T., & Baars, B. J. (1976a). Semantic bias effects on the outcomes of verbal slips. *Cognition, 4*, 177-187.

Mowrey, R. A., & MacKay, I. R. (1990). Phonological primitives: Electromyographic speech error evidence. *Journal of the Acoustical Society of America, 88*, 1299-1312.

Murray, L. L. (2000). The effects of varying attentional demands on the word retrieval skills of adults with aphasia, right hemisphere brain damage, or no brain damage. *Brain and Language, 72*(1), 40-72.

Nespoulous, J.-L., Joanette, Y., Beland, R., Caplan, D., & Lecours, A. R. (1984). Phonologic disturbances in aphasia: Is there a "markedness effect" in aphasic phonetic errors? In F. C. Rose (Ed.), *Advances in Neurology: Progress in Aphasiology* (Vol. 42, pp. 203-214). New York, NY: Raven Press.

Newman, R., Sawusch, J., & Luce, P. (1999). Underspecification and phoneme frequency in speech perception. In M. Broc & J. Pierrehumbert (Eds.), *Papers in Laboratory Phonology V: Language Acquisition and the Lexicon* (pp. 298-311). Cambridge, UK: Cambridge University Press.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighbourhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*(3), 873-889.

Nickels, L. (1997). *Spoken Word Production and its Breakdown in Aphasia*. East Sussex: Psychology Press.

Nickels, L., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex, 31*, 209-237.

Niemi, J., Koivuselka-Sallinen, P., & Laine, M. (1987). Lexical deformations are sensitive to morphosyntactic factors in posterior aphasics. *Aphasiology, 1*(1), 53-57.

Nooteboom, S. G. (1973). The tongue slips into patterns. In V. Fromkin (Ed.), *Speech Errors as Linguistic Evidence*. The Hague: Mouton.

Nooteboom, S. G. (1980). Ch. 6: Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In V. A. Fromkin (Ed.), *Errors in*

Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand (pp. 87-95). New York, NY: Academic Press.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273-281.

Parsons, C. L., Lambier, J. D., & Miller, A. (1988). Phonological processes and phonemic paraphasias. *Aphasiology, 2*(1), 45-54.

Pate, D. S., Saffran, E. M., & Martin, N. (1987). Specifying the nature of the production impairment in a conduction aphasic: A case study. *Language and Cognitive Processes, 2*(1), 43-84.

Pease, D. M., & Goodglass, H. (1978). The effects of cuing on picture naming in aphasia. *Cortex, 14*, 178-189.

Pick, A. (1931). *Aphasia* (Brown, J. W. (1973), Trans.). Springfield, IL: Charles C. Thomas.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, (in press).*

Roach, A., Schwartz, M. F., Martin, N., Grewal, R. S., & Brecher, A. (1996). The Philadelphia Naming Test: Scoring and rationale. *Clinical Aphasiology, 24*, 121-133.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107-142.

Romani, C., & Calabrese, A. (1998). Syllabic constraints in the phonological errors of an aphasic patient. *Brain and Language, 64*(1), 83-121.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24*, 210-231.

Savage, G. R., Bradley, D. C., & Forster, K. I. (1990). Word frequency and the pronunciation task: The contribution of articulatory fluency. *Language and Cognitive Processes, 5*, 203-236.

Schlenk, K.-J., Huber, W., & Willmes, K. (1987). 'Prepairs' and repairs: Different monitoring functions in aphasic language production. *Brain and Language, 30*, 226-244.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86-102.

Schwartz, M. F. (1984). What the classical aphasia categories can't do for us, and why. *Brain and Language, 21*, 3-8.

Schwartz, M. F., & Brecher, A. (2000). A model-driven analysis of severity, response characteristics, and partial recovery in aphasics' picture naming. *Brain and Language, 73*(1), 62-91.

Schwartz, M. F., Saffran, E. M., Bloch, D. E., & Dell, G. S. (1994). Disordered speech production in aphasic and normal speakers. *Brain and Language, 47*(1), 52-88.

Sears, C. R., Hino, Y., & Lupker, S. J. (1995). Neighborhood size and neighborhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*(4), 876-900.

Segui, J., & Grainger, J. (1990). Priming word recognition with orthographic neighbors: Effects of relative prime-target frequency. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 65-76.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53*, 91-127.

Sevald, C. A., Dell, G. S., & Cole, J. S. (1995). Syllable structure in speech production: Are syllables chunks or schemas? *Journal of Memory and Language, 34*, 807-820.

Shattuck, S. R. (1975). *Speech Errors and Sentence Production.* Unpublished doctoral dissertation, MIT, Massachusetts.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial ordering mechanism in speech production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence Processing.* Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 109-136). New York, NY: Springer-Verlag.

Shattuck-Hufnagel, S. (1987). The role of word-onset consonants in speech production planning. In E. Keller & M. Gopnik (Eds.), *Motor and Sensory Processing in Language*. Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42*, 213-259.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech error production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18*, 41-55.

Shattuck-Hufnagel, S., & Klatt, D. H. (1980). How single phoneme error data rule out two models of error generation. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand* (pp. 35-46). New York, NY: Academic Press.

Sheridan, R. B. (1906). *The Rivals.* Boston.

Shinn, P., & Blumstein, S. E. (1983). Phonetic disintegration in aphasia: Acoustic analysis of spectral characteristics for place of articulation. *Brain and Language, 20*, 90-114.

Slowiaczek, L. M., & Hamburger, M. B. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(6), 1239-1250.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Phonological priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 64-75.

Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition, 14*(3), 230-237.

Soloman, R. L., & Postman, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *Journal of Experimental Psychology, 43*, 195-201.

Spencer, K. A., Doyle, P. J., McNeil, M. R., Wambaugh, J. L., Park, G., & Carroll, B. (2000). Examining the facilitative effects of rhyme in a patient with output lexicon damage. *Aphasiology, 14*(5/6), 567-584.

Stemberger, J. P. (1982b). The nature of segments in the lexicon: Evidence from speech errors. *Lingua, 56*, 235-259.

Stemberger, J. P. (1984a). Structural errors in normal and agrammatic speech. *Cognitive Neuropsychology, 1*, 281-313.

Stemberger, J. P. (1985). *The Lexicon in a Model of Language Production*. New York, NY: Garland Publishing.

Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition, 35*, 123-157.

Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language, 30*, 161-185.

Stemberger, J. P. (1992). Ch. 8: The reliability and replicability of naturalistic speech error data: A comparison with experimentally induced errors. In B. J. Baars (Ed.), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (pp. 195-215). New York, NY: Plenum Press.

Stemberger, J. P., & MacWhinney, B. (1986). Frequency and the lexical storage of regularly inflected forms. *Memory and Cognition, 14*(1), 17-26.

Stemberger, J. P., & Treiman, R. (1986). The internal structure of word-initial consonant clusters. *Journal of Memory and Language, 25*, 163-180.

Sussman, H. M. (1984). A neuronal model for syllable representation. *Brain and Language, 22*, 167-177.

Svartvik, J., & Quirk, R. (1980). *A Corpus of English Conversation*. Lund: Gleerup.

Talo, E. S. (1980). Ch. 5: Slips of the tongue in normal and pathological speech. In V. A. Fromkin (Ed.), *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen, and Hand* (pp. 81-86). New York, NY: Academic Press.

Taraban, R., & McClelland, J. L. (1987). Conspiracy effects in word pronunciation. *Journal of Memory and Language, 26*, 608-631.

Tent, J., & Clark, J. E. (1980). An experimental investigation into the perception of slips of the tongue. *Journal of Phonetics, 8*, 317-325.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15*, 49-74.

Trost, J. E., & Canter, G. J. (1974). Apraxia of speech in patients with Broca's aphasia: A study of phoneme production accuracy and error patterns. *Brain and Language, 1*, 63-80.

Tuller, B. (1984). On categorizing aphasic speech errors. *Neuropsychologia, 22*(5), 547-557.

Valdois, Joanette, Y., & Nespoulos, J.-L. (1989). Intrinsic organization of sequences of phonemic approximations: A preliminary study. *Aphasiology, 3*, 55-73.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech, 40*(3), 211-228.

Vitevitch, M. S. (ms in prep). The influence of phonological similarity neighbourhoods on speech production.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325-329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics in spoken word recognition. *Journal of Memory and Language, 40*, 374-408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40*(1), 47-62.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*, 306-311.

Vitevitch, M. S., & Sommers, M. S. (ms in prep). The role of neighbors in the tip-of-the-tongue state.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*, 392-393.

Wheeler, D. W., & Touretzky, D. S. (1997). A parallel licensing model of normal slips and phonemic paraphasias. *Brain and Language, 59*(1), 147-201.

Williams, S. E., & Canter, G. J. (1982). The influence of situational context on naming performance in aphasic syndromes. *Brain and Language, 17*, 92-106.

Wright, J. F., & Ahmad, K. (1997). The connectionist simulation of aphasic naming. *Brain and Language, 59*(2), 367-389.

Zingeser, L. B., & Berndt, R. S. (1990). Retrieval of nouns and verbs in agrammatics vs anomics. *Brain and Language, 39*(1), 14-32.

Zwicky, A. (1982). Classical malapropisms and the creation of a mental lexicon. In L. K. Obler & L. Menn (Eds.), *Exceptional Language and Linguistics*. San Diego, CA: Academic Press.

Figure 4-i. Pilot study: Proportions of errors for each subject (phonological vs other errors)

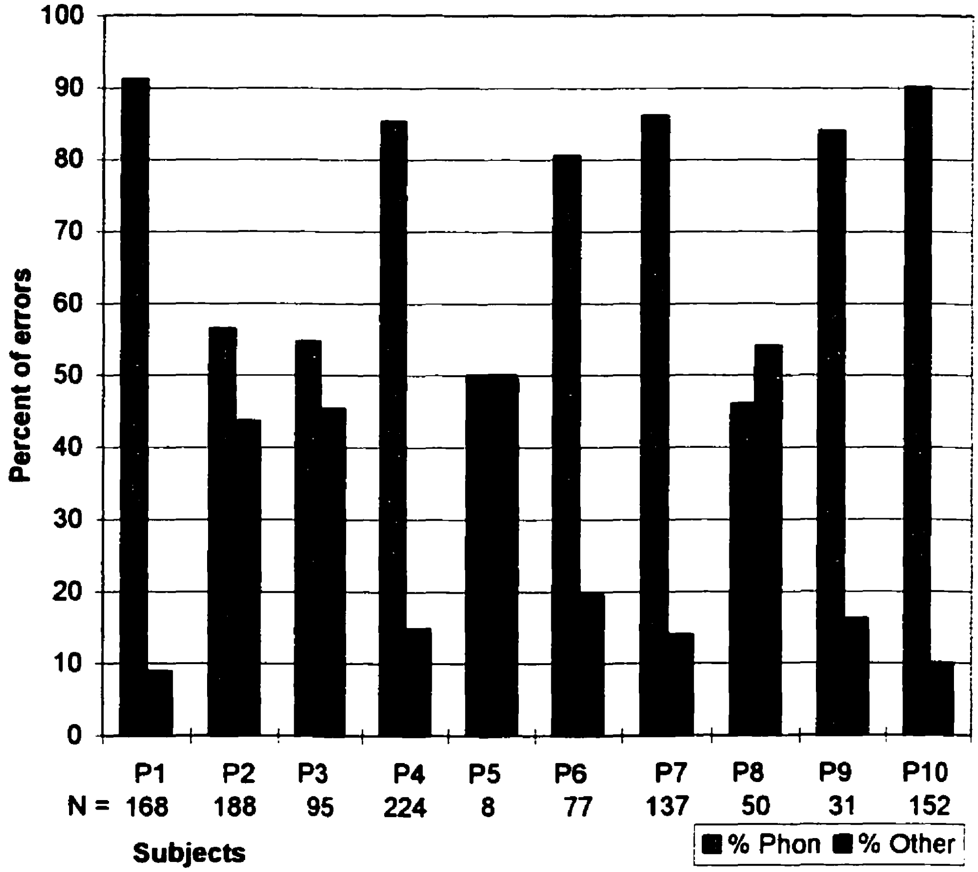


Figure 4-ii. Pilot study: Proportions of contextual error types (group pattern)

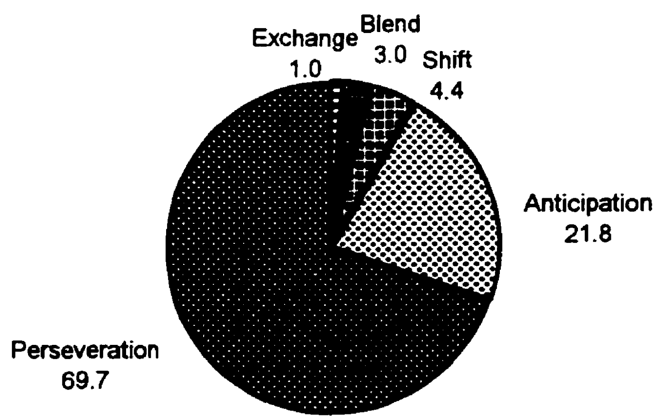


Figure 4-iii. Pilot study: Proportions of contextual error types for individual subjects showing A:P ratios consistent with the group pattern

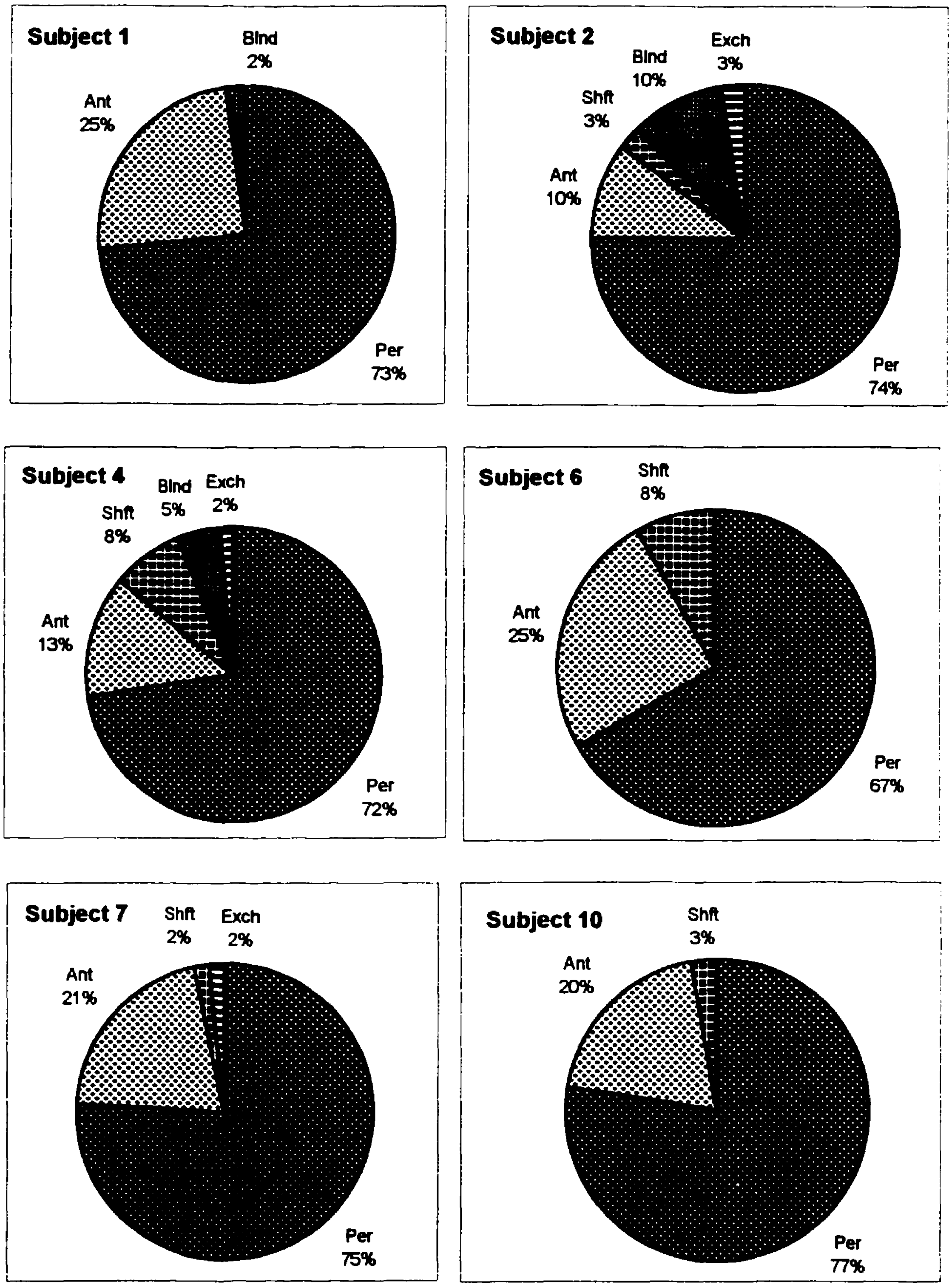


Figure 4-iv. Pilot study: Proportions of contextual error types for individual subjects showing A:P ratios inconsistent with the group pattern

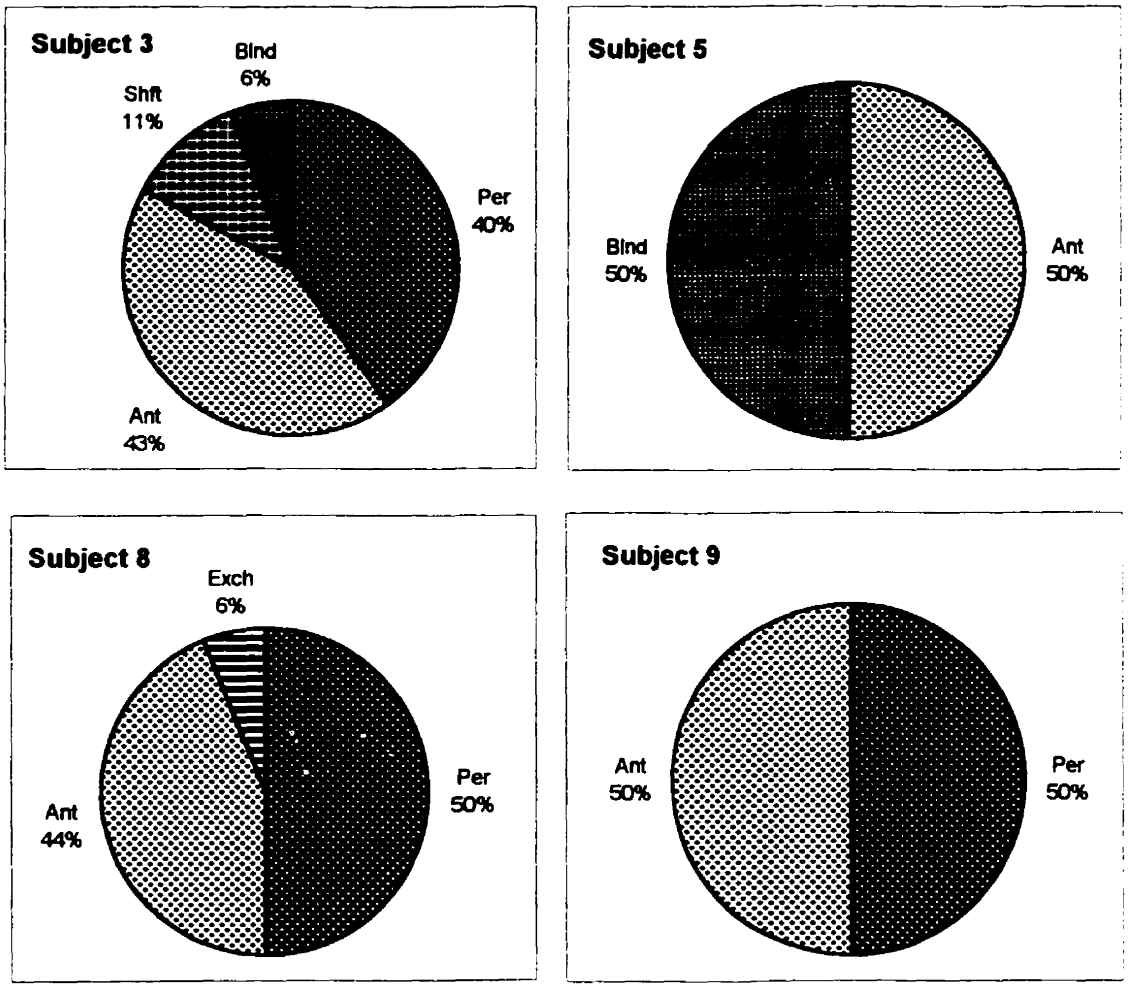
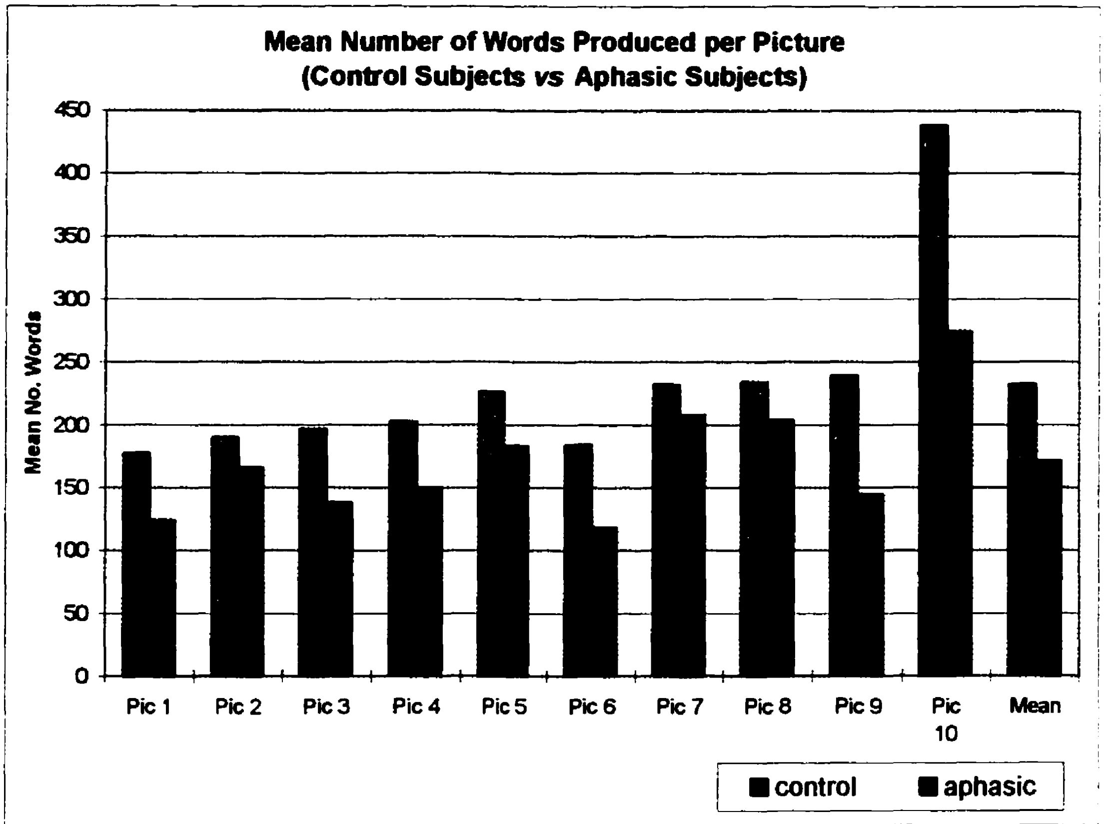


Figure 5-i. NR task: Mean word counts and phonological error incidence for each picture



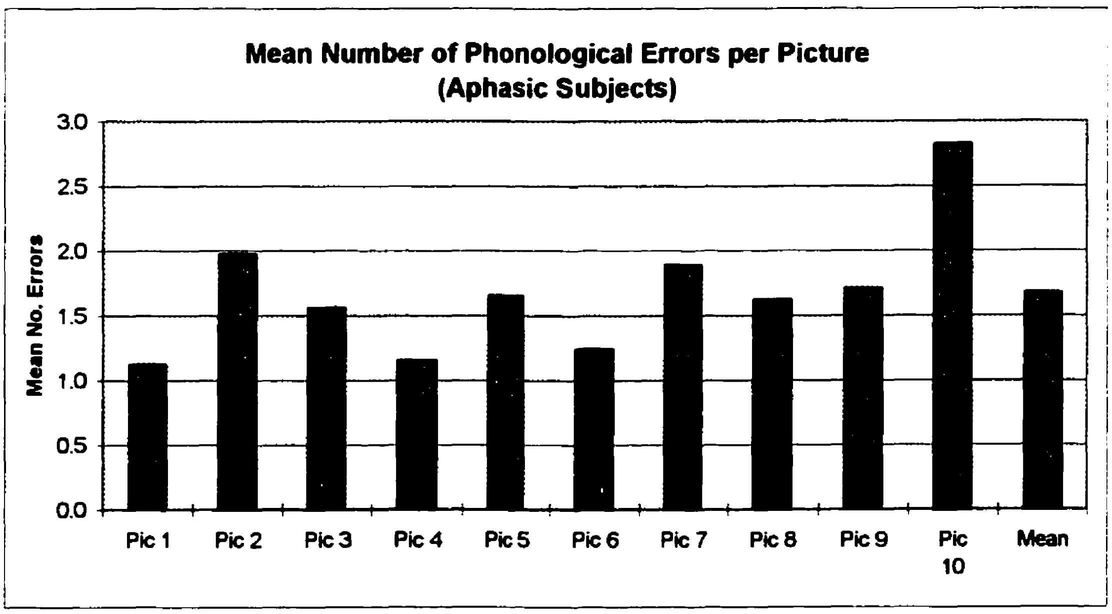
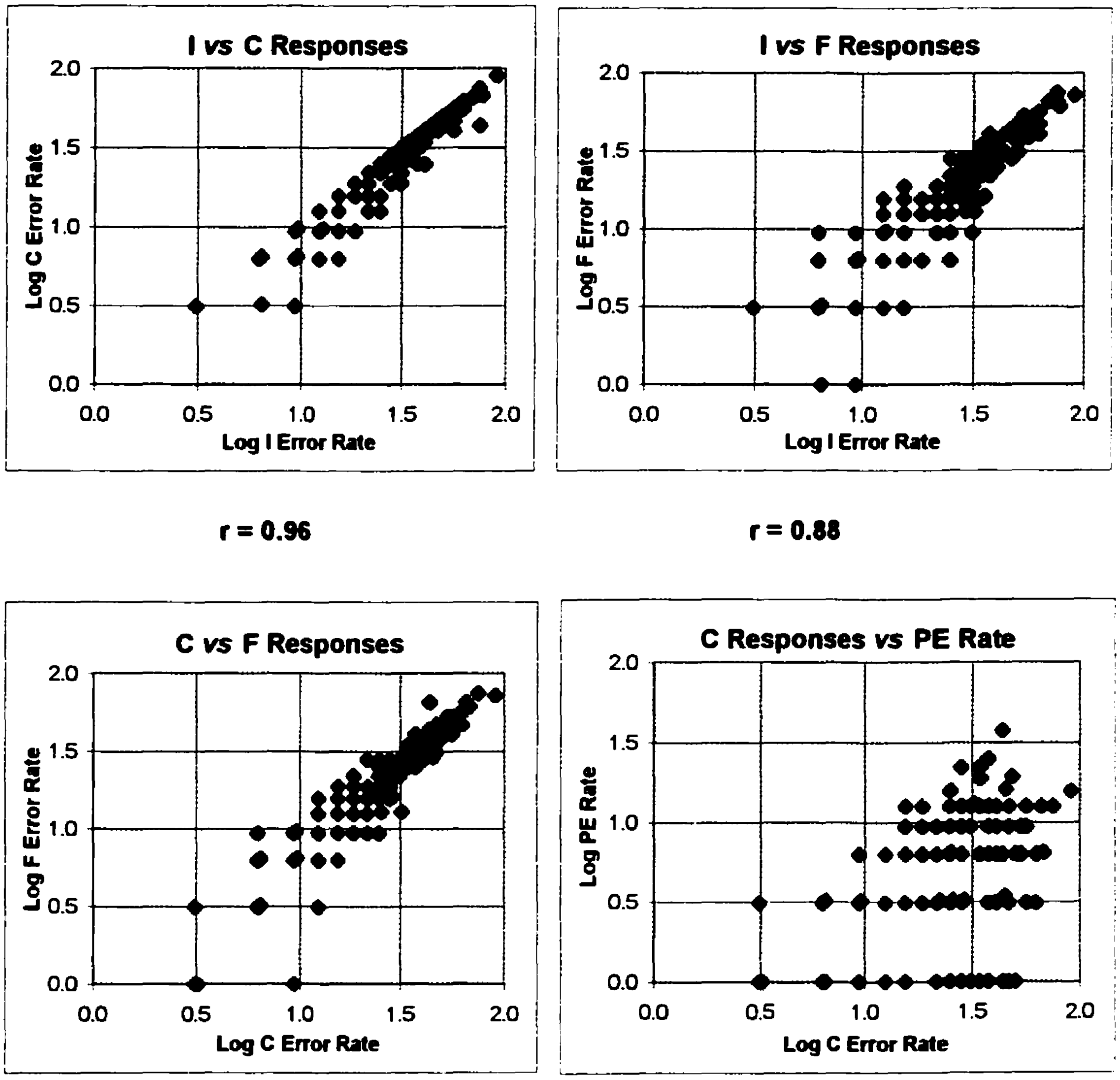


Figure 5-ii. Correlations among PNT accuracy measures (total errors on initial (I), complete (C), and final (F) responses, and phonological error (PE) rate)



r = 0.96

r = 0.88

r = 0.92

r = 0.52

Figure 5-iii. NR task: Mean values for item and neighbourhood variables (error-targets vs control-targets)

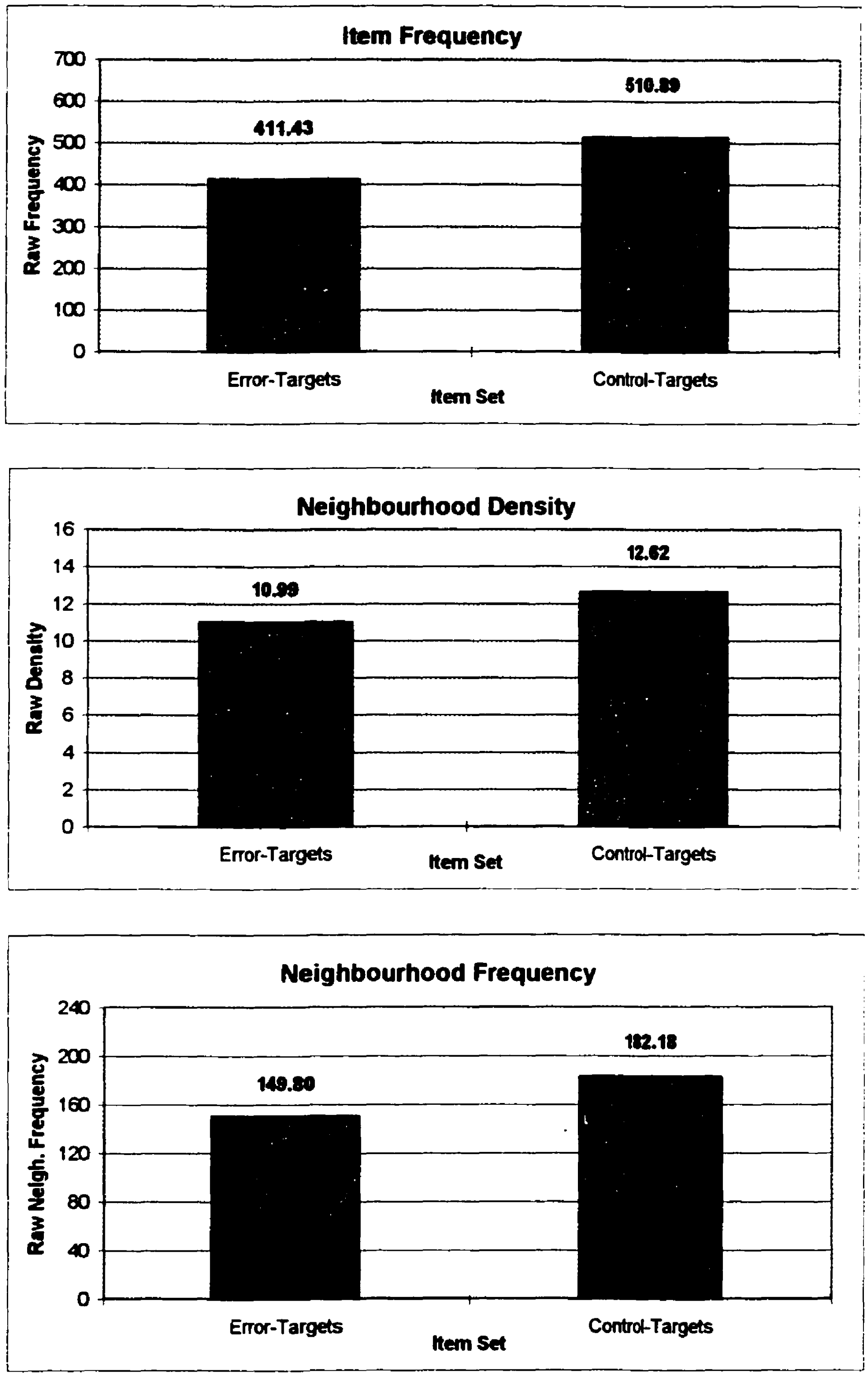


Figure 5-iv. NR task: Mean values for item and neighbourhood variables (errors vs targets; words vs non-words)

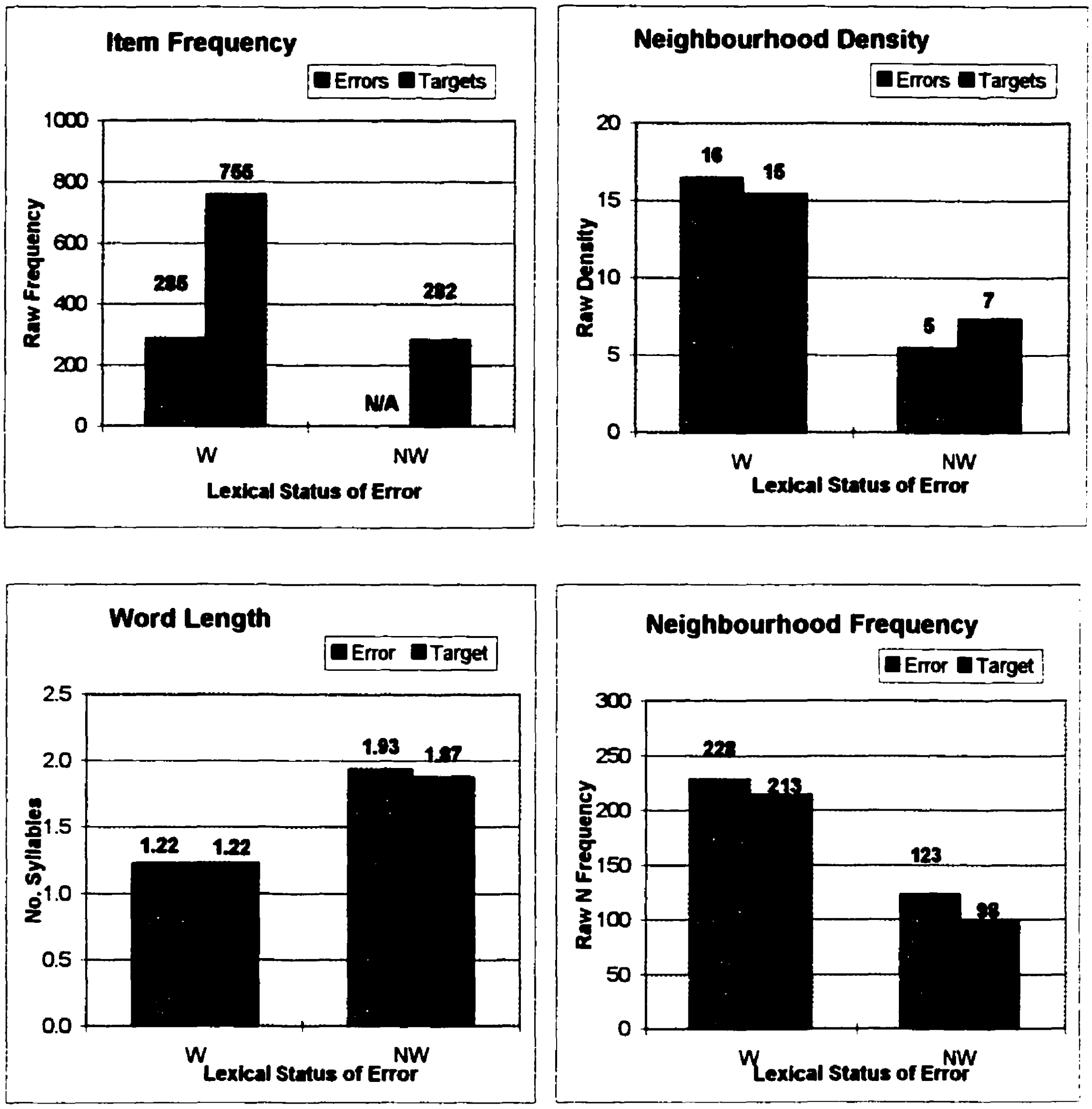
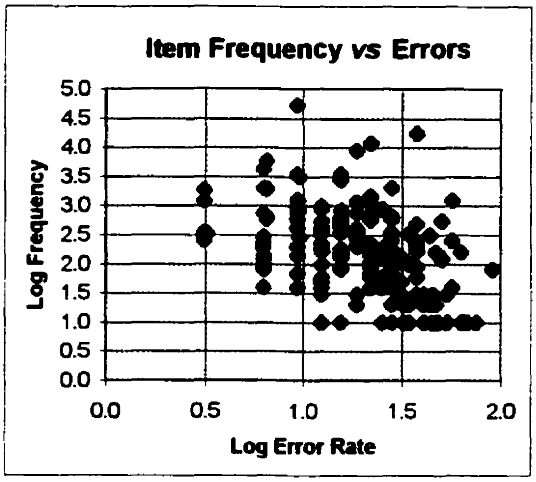
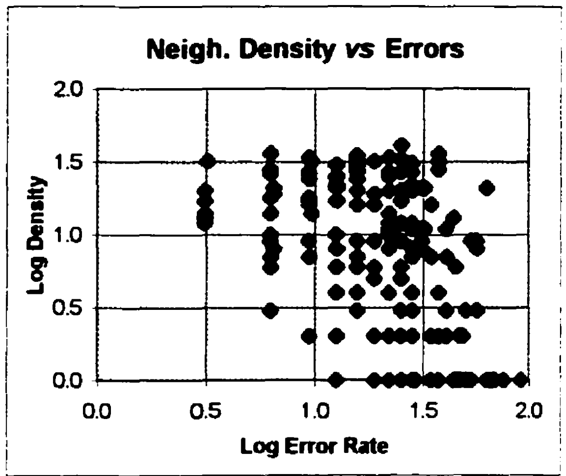


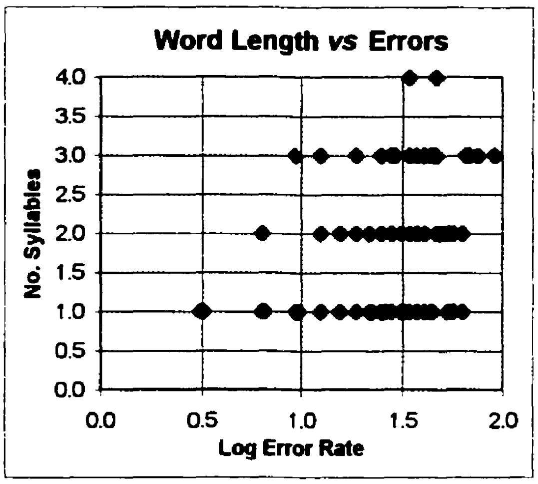
Figure 5-v.a) Correlations of PNT accuracy (first complete response) with item and neighbourhood variables



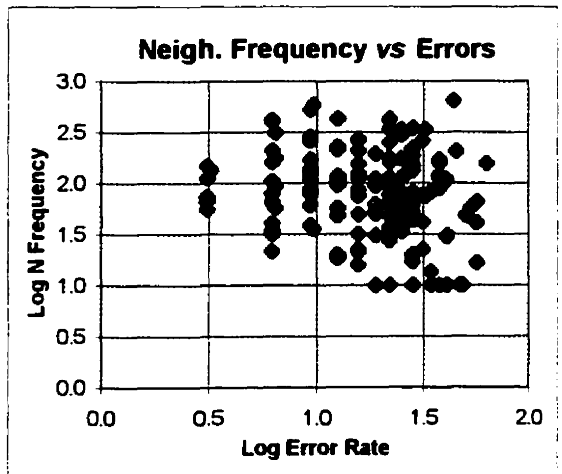
r = - 0.44



r = - 0.41

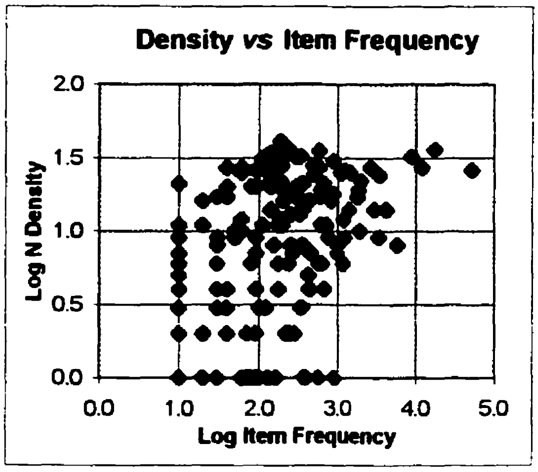


r = 0.40

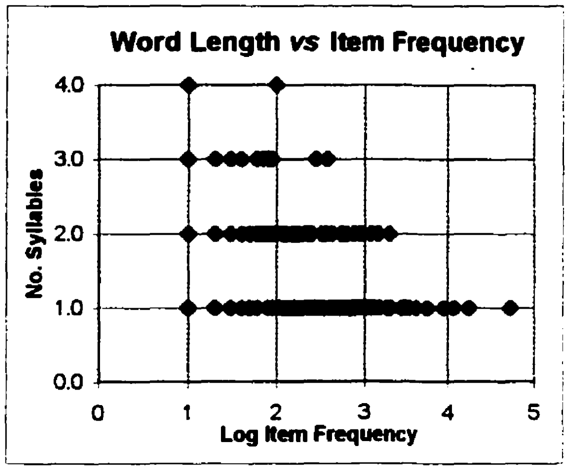


r = - 0.18

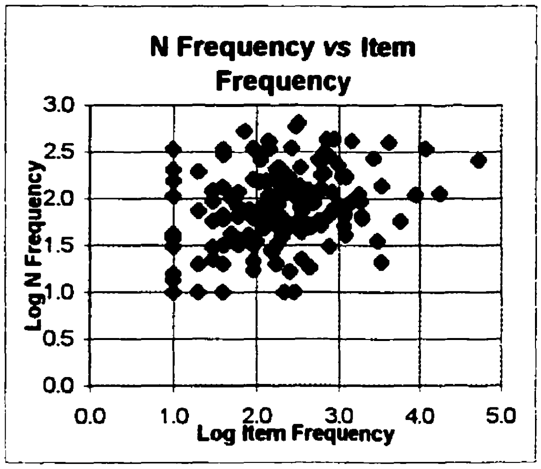
Figure 5-v.b) Correlations among item and neighbourhood variables (PNT items)



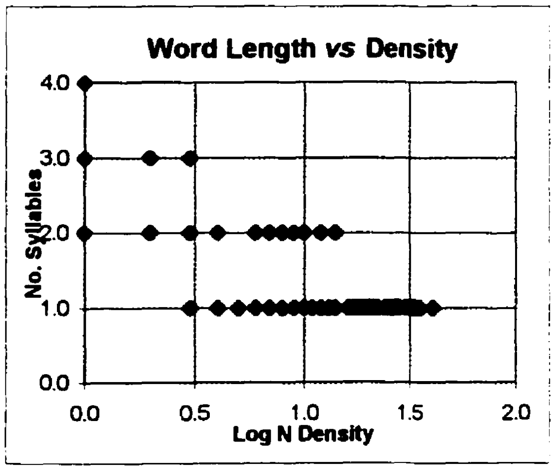
r = 0.41



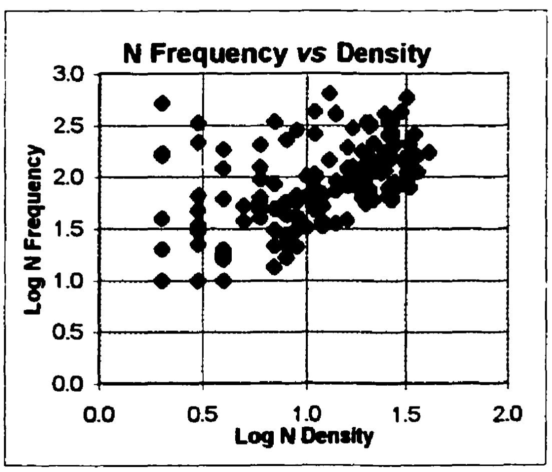
r = - 0.36



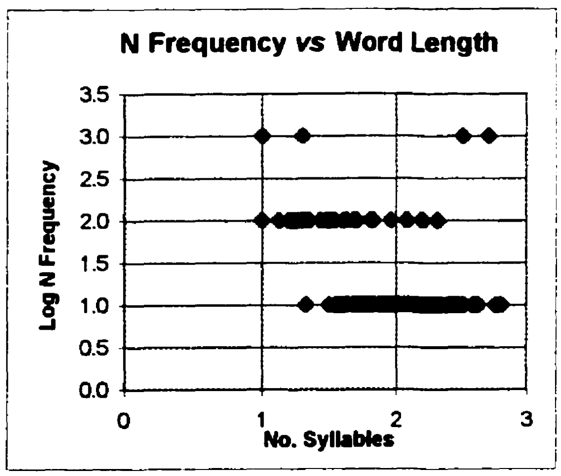
r = 0.37



r = - 0.76



r = 0.61



r = - 0.56

Figure 5-vi. PNT task: Mean values for item and neighbourhood variables (errors vs targets; words vs non-words)

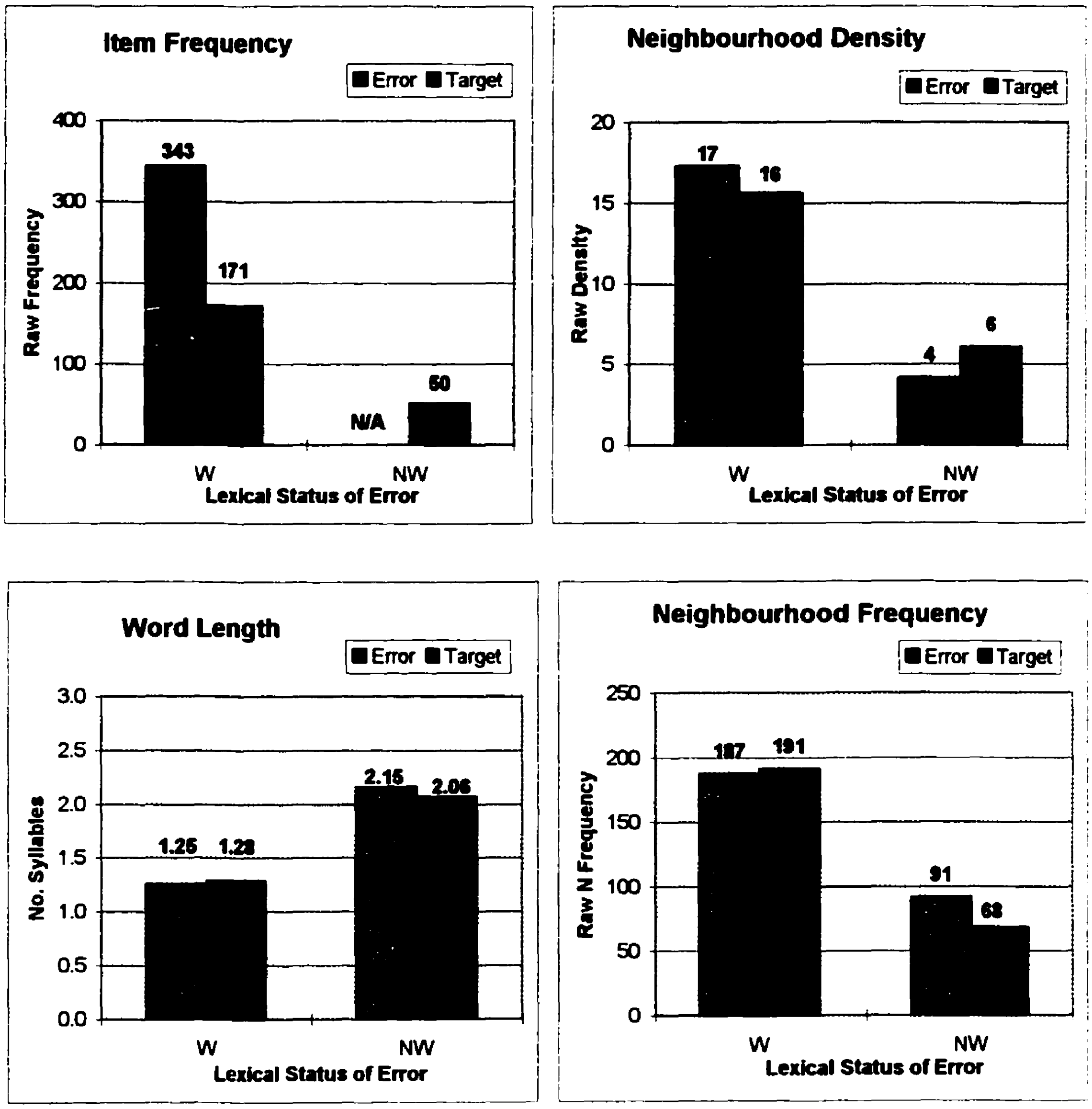
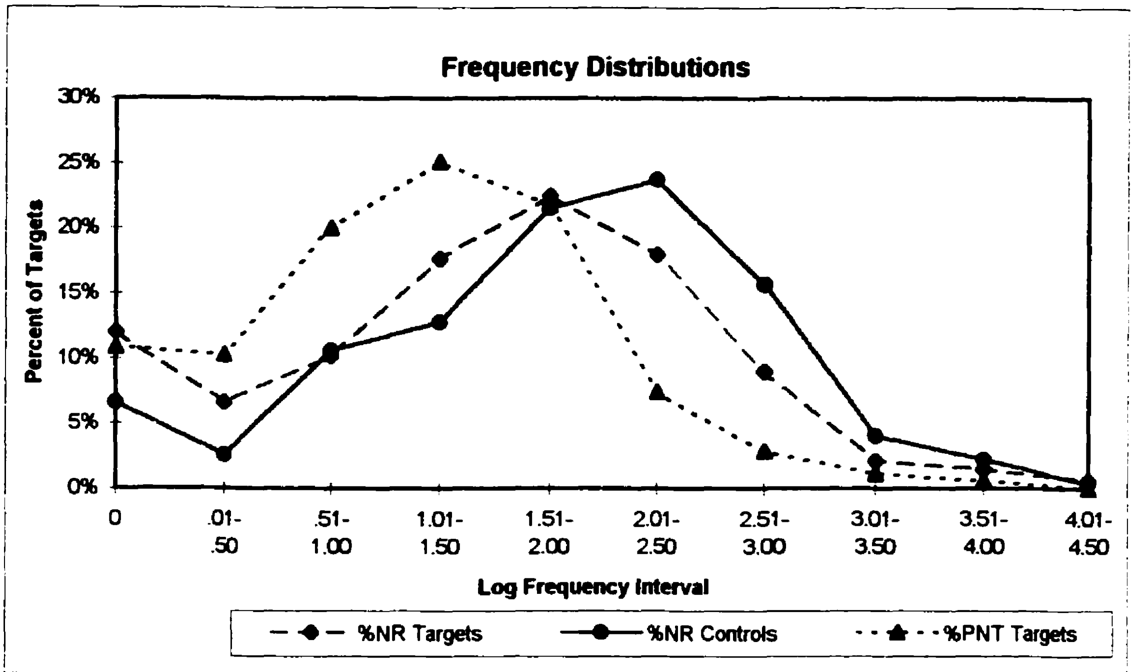


Figure 5-vii. Distributions of target items across corpora
a) Item characteristics



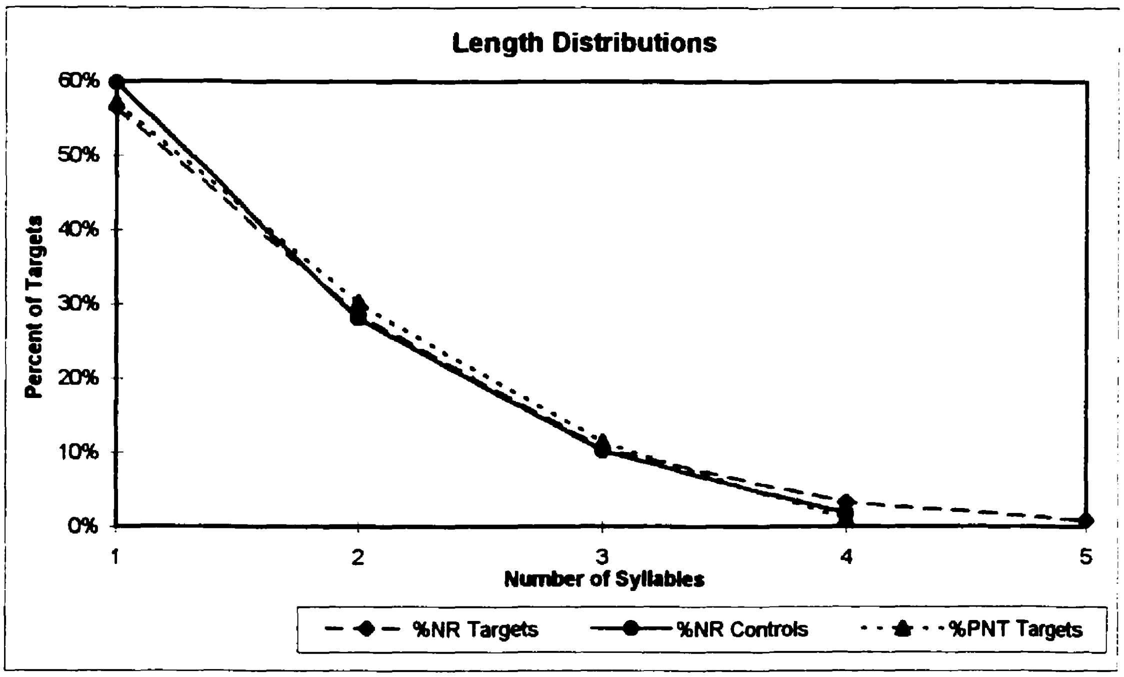
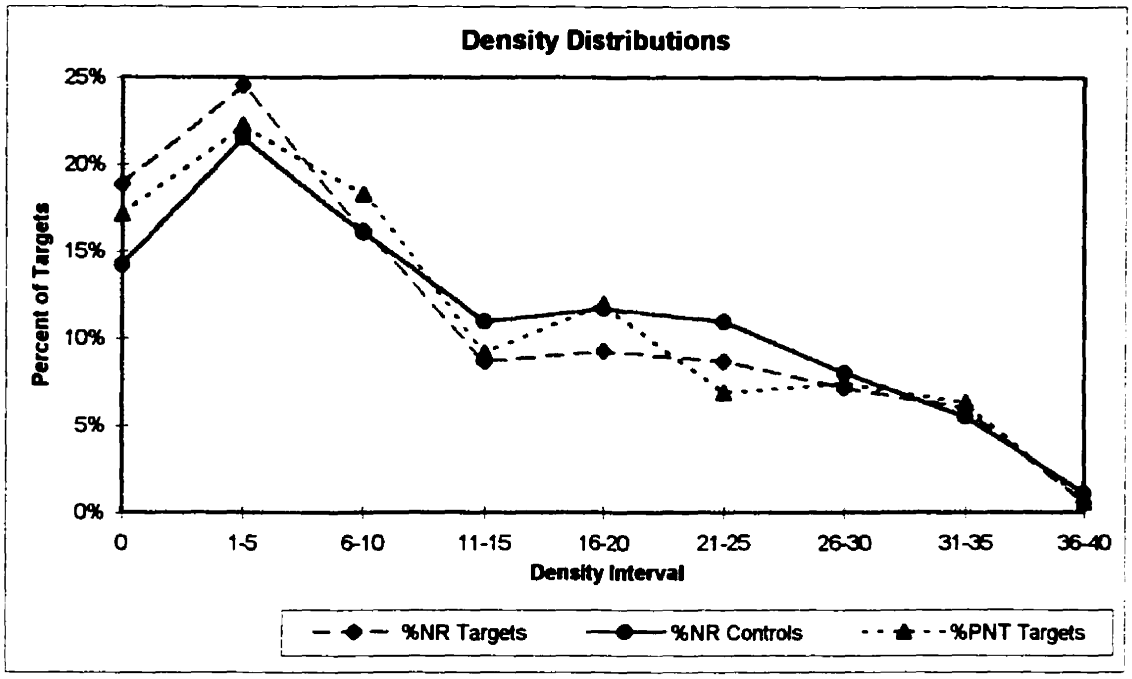


Figure 5-vii. Distributions of target items across corpora
b) Neighbourhood characteristics



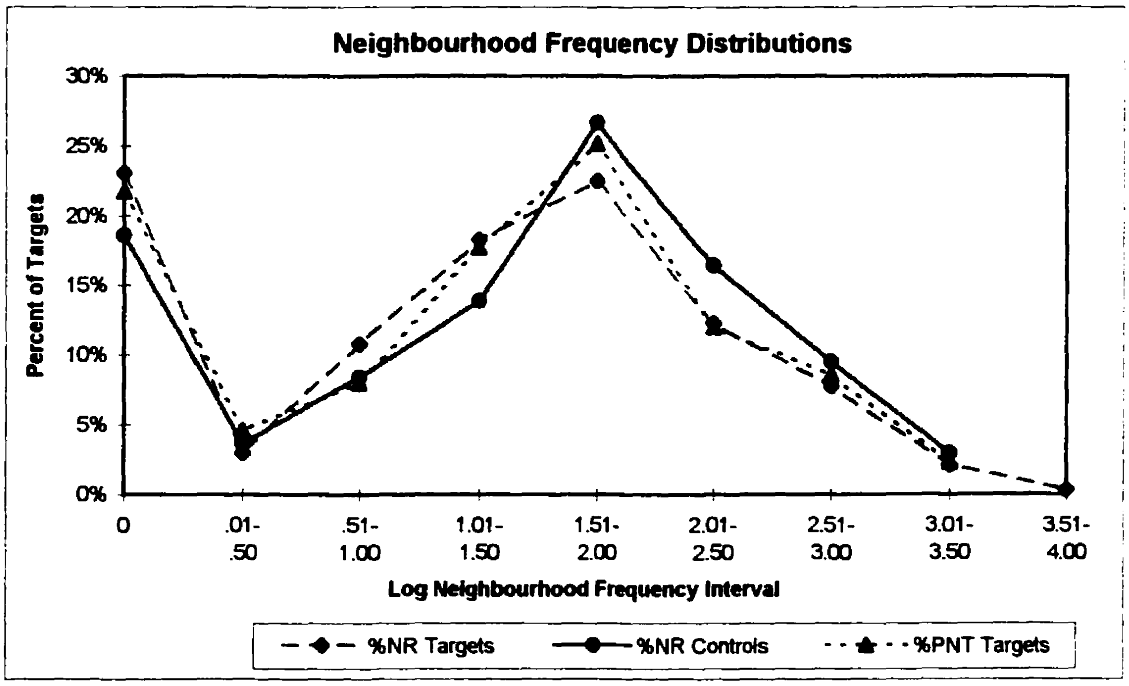
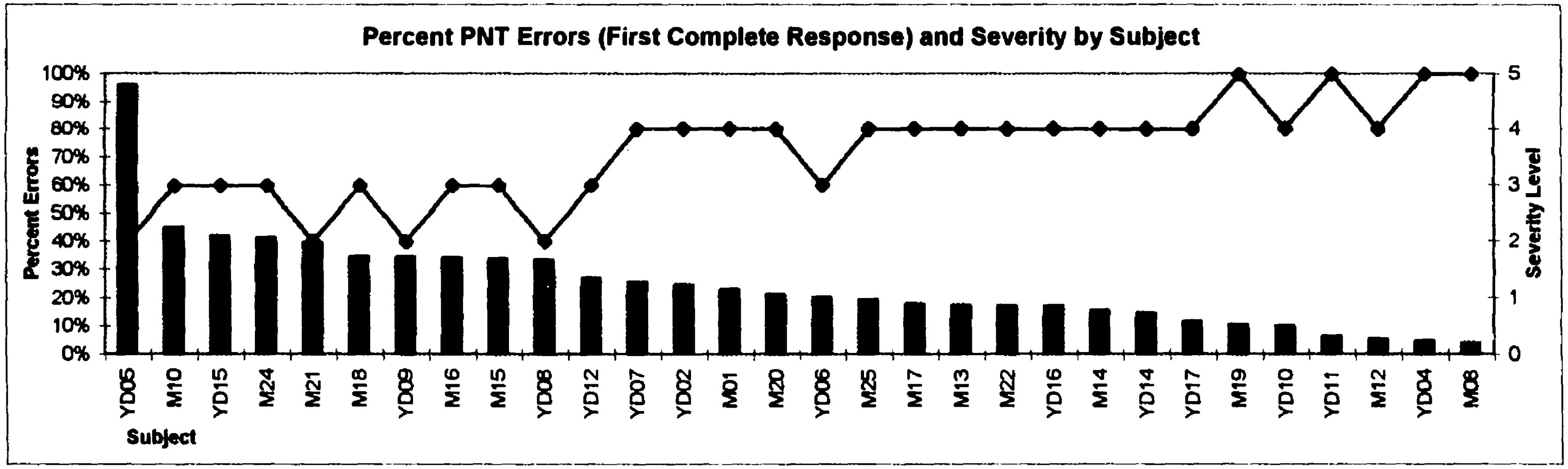


Figure 5-viii. Error distributions and severity level by subject



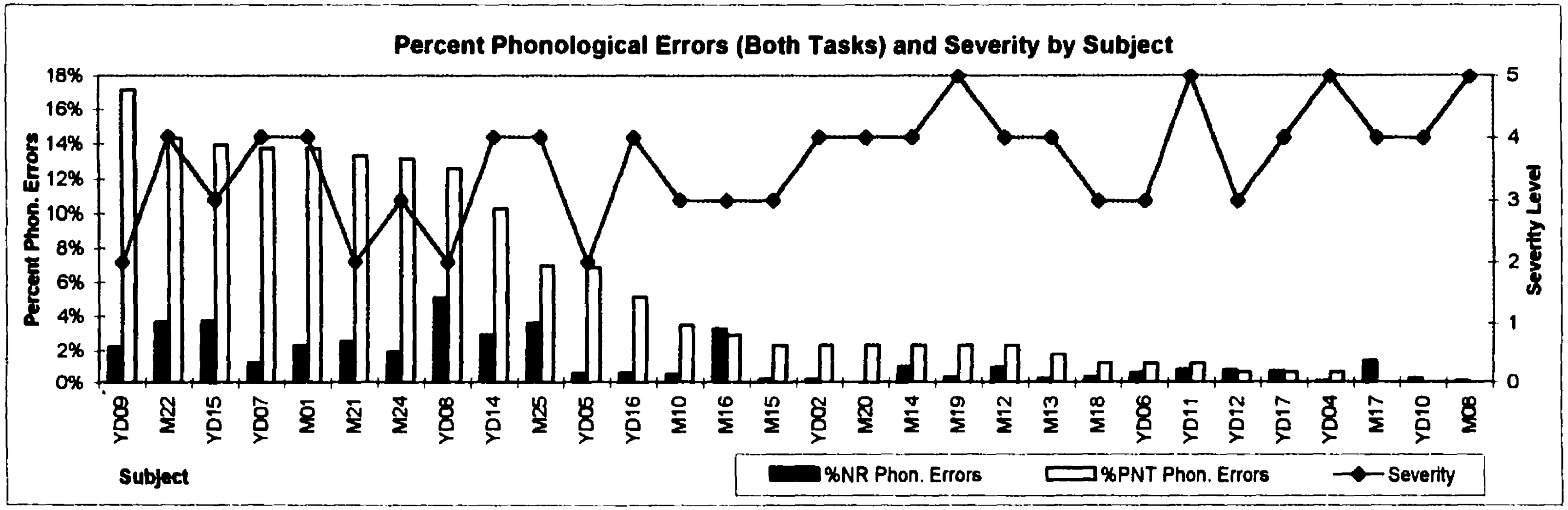
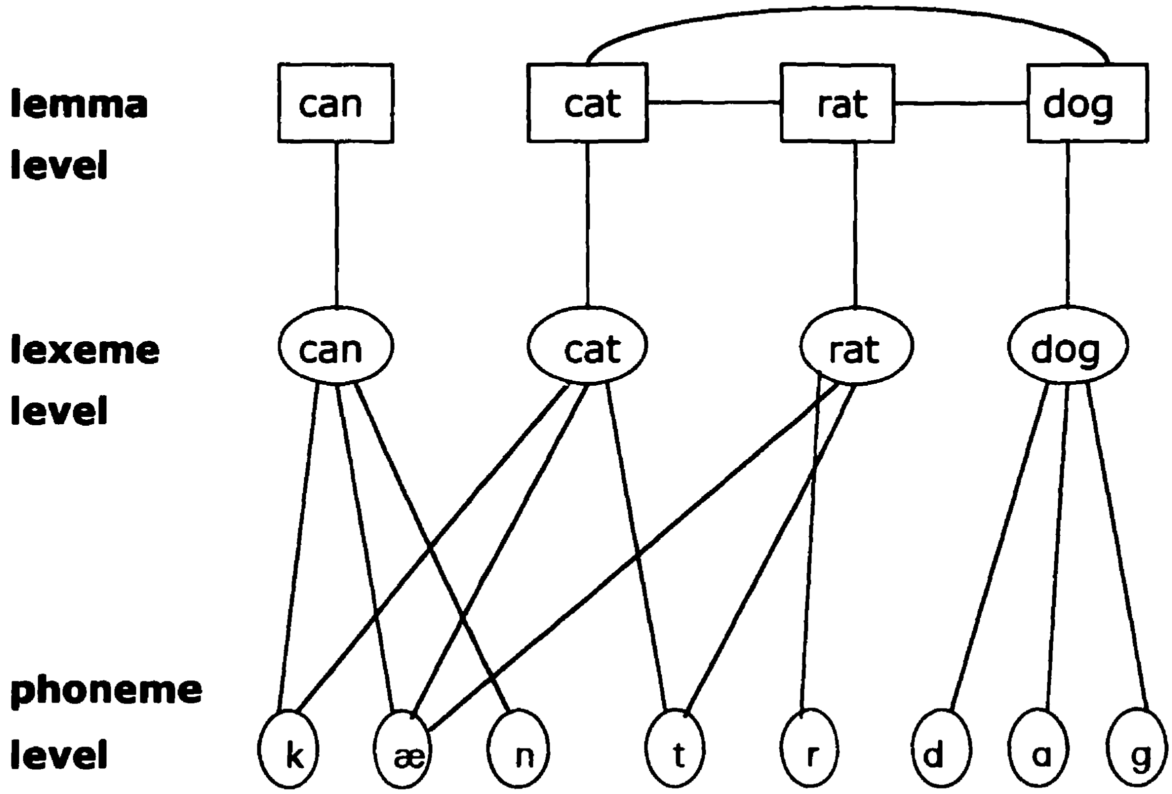


Figure 6-i. Interactive activation model of speech production



(Adapted from Dell, 1986)

Appendix 4-i. BDAE Stimuli Used in Pilot Study

Confrontation Naming Task

Objects	Actions	Body Parts
chair	running	ear
key	sleeping	nose
glove	drinking	elbow
feather	smoking	shoulder
hammock	falling	ankle
cactus	dripping	wrist

Sentence Repetition Task

High-Frequency Sentences

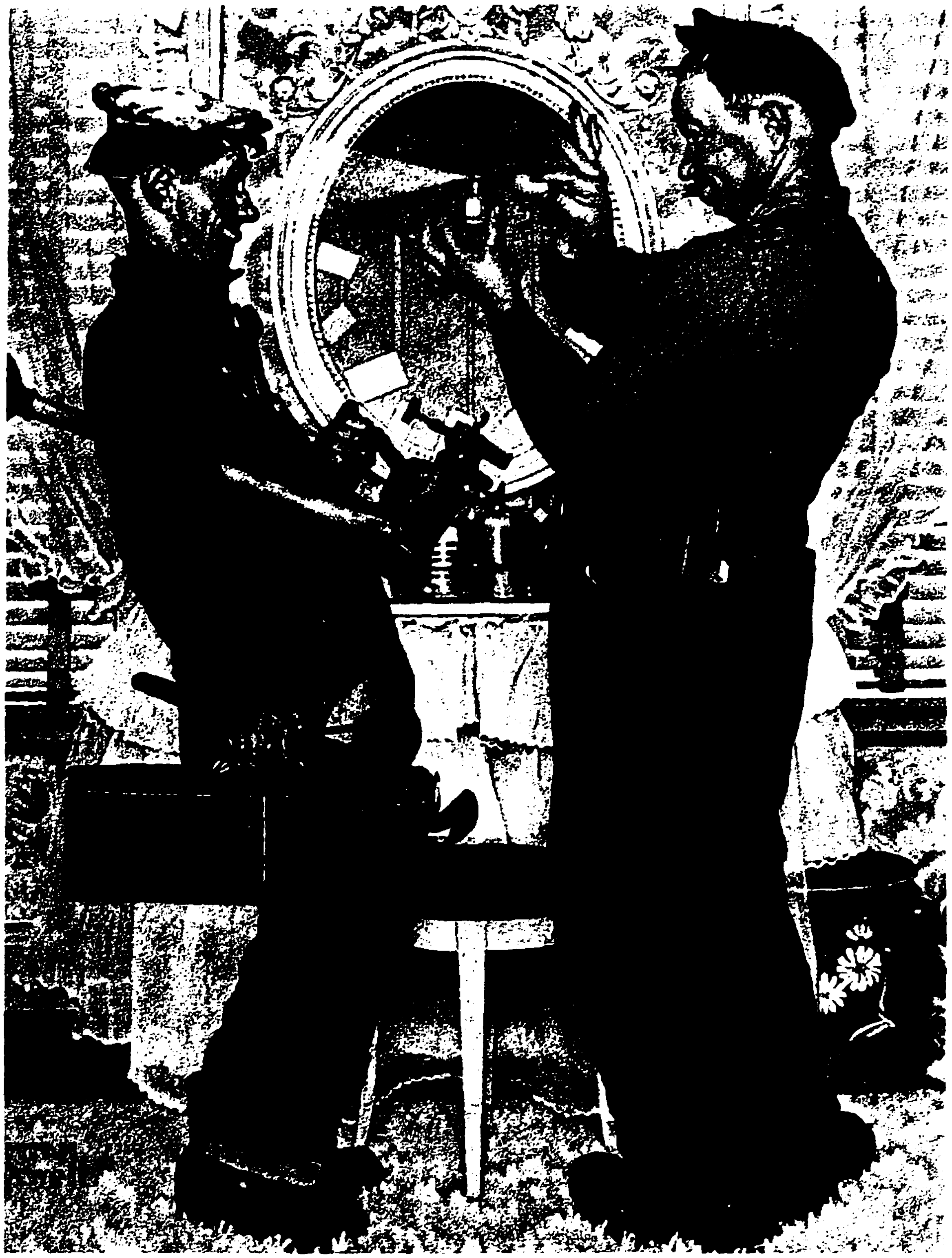
a. You know how.
b. Down to earth.
c. I got home from work.
d. You should not tell her.
e. Go ahead and do it if possible.
f. Near the table in the dining room.
g. They heard him speak on the radio last night.
h. I stopped at his front door and rang the bell.

Low-Frequency Sentences

a. The vat leaks.
b. Limes are sour.
c. The spy fled to Greece.
d. Pry the tin lid off.
e. The Chinese fan had a rare emerald.
f. The barn swallow captured a plump worm.
g. The lawyer's closing argument convinced him.
h. The phantom soared across the foggy heath.

Appendix 5-i. Norman Rockwell Pictures

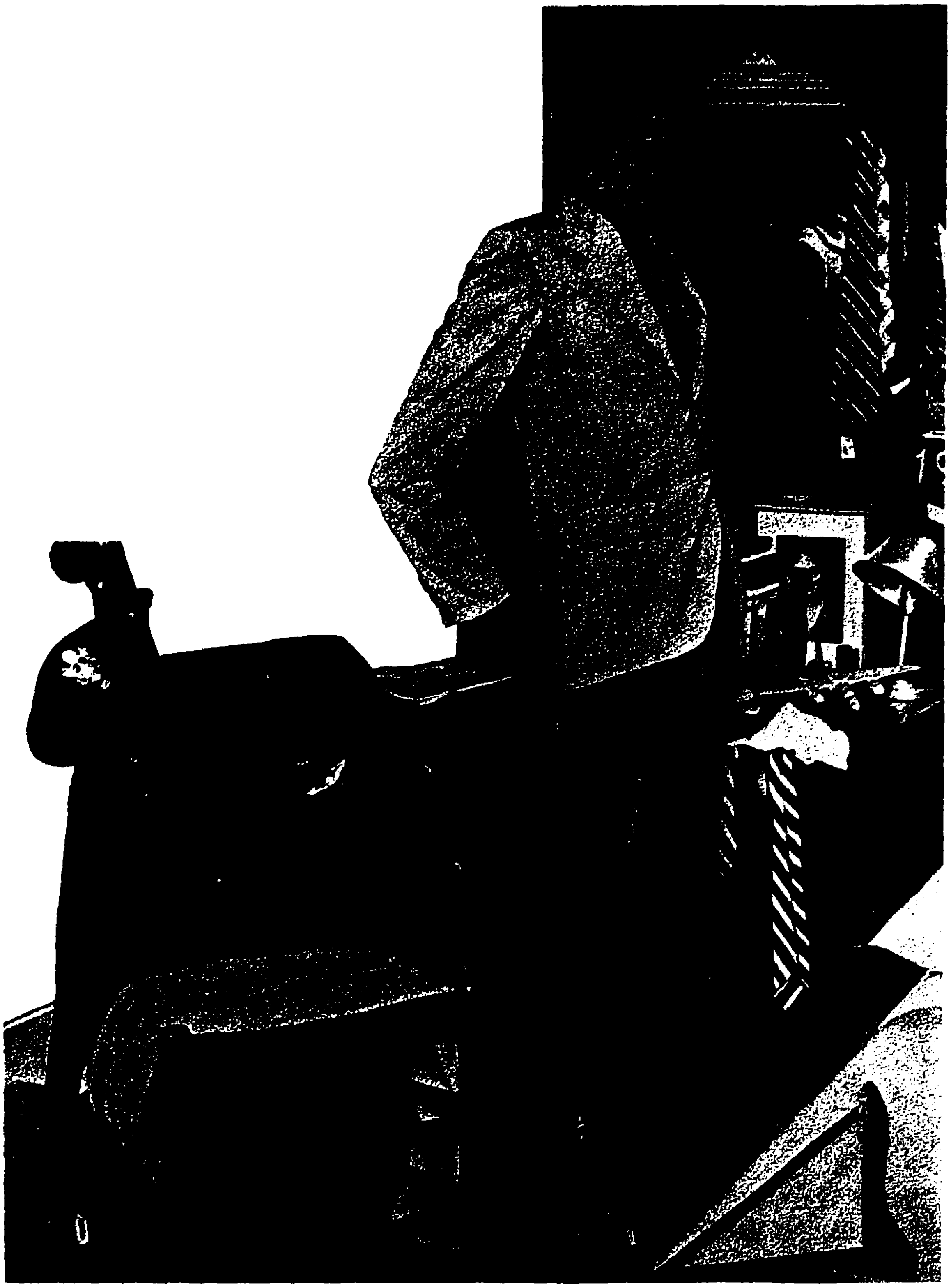
Picture # 1: Doctor's Office



Picture # 2: The Plumbers

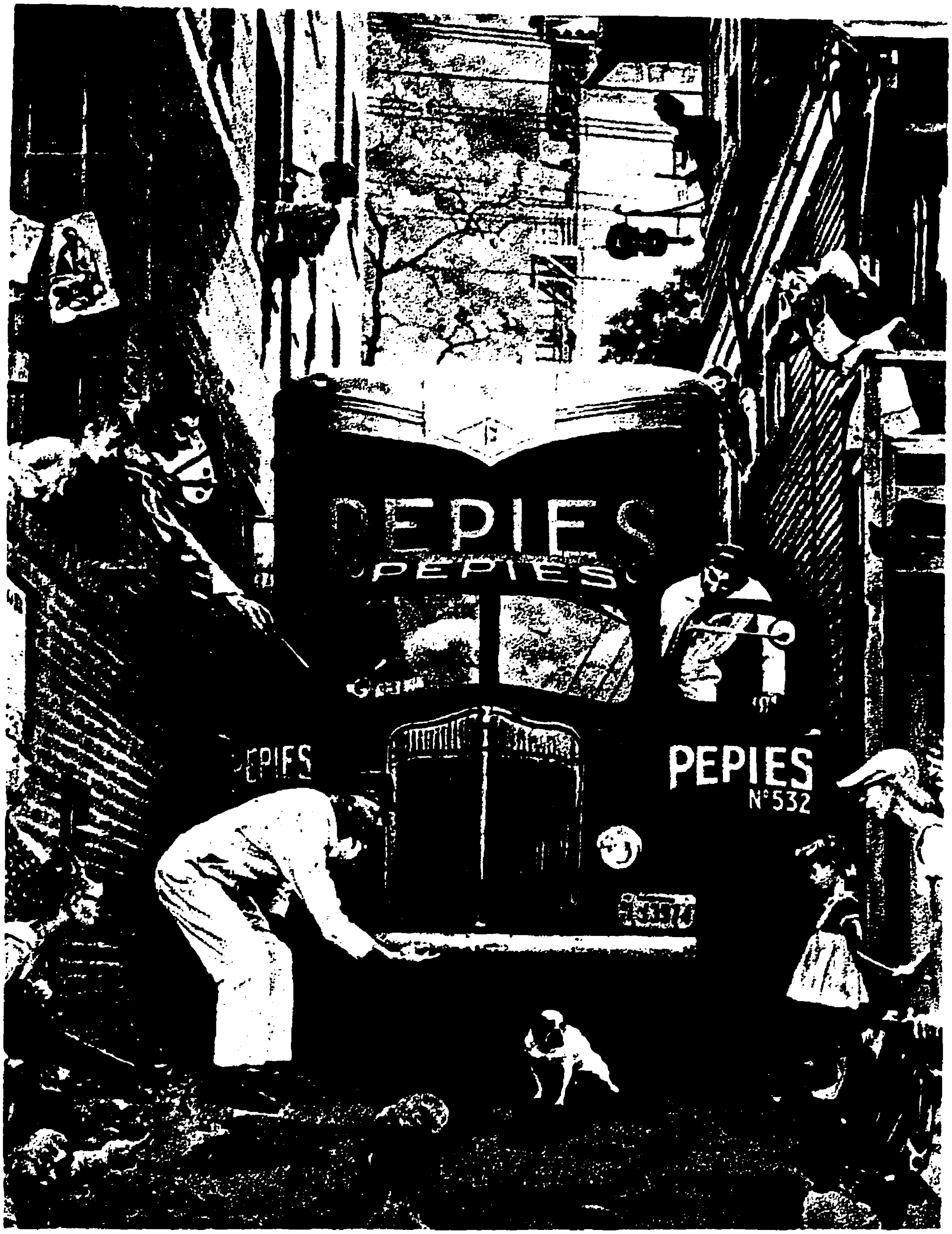
Picture # 3: The Window Washer

Picture # 4: Easter Morning



Picture # 5: An Imperfect Fit

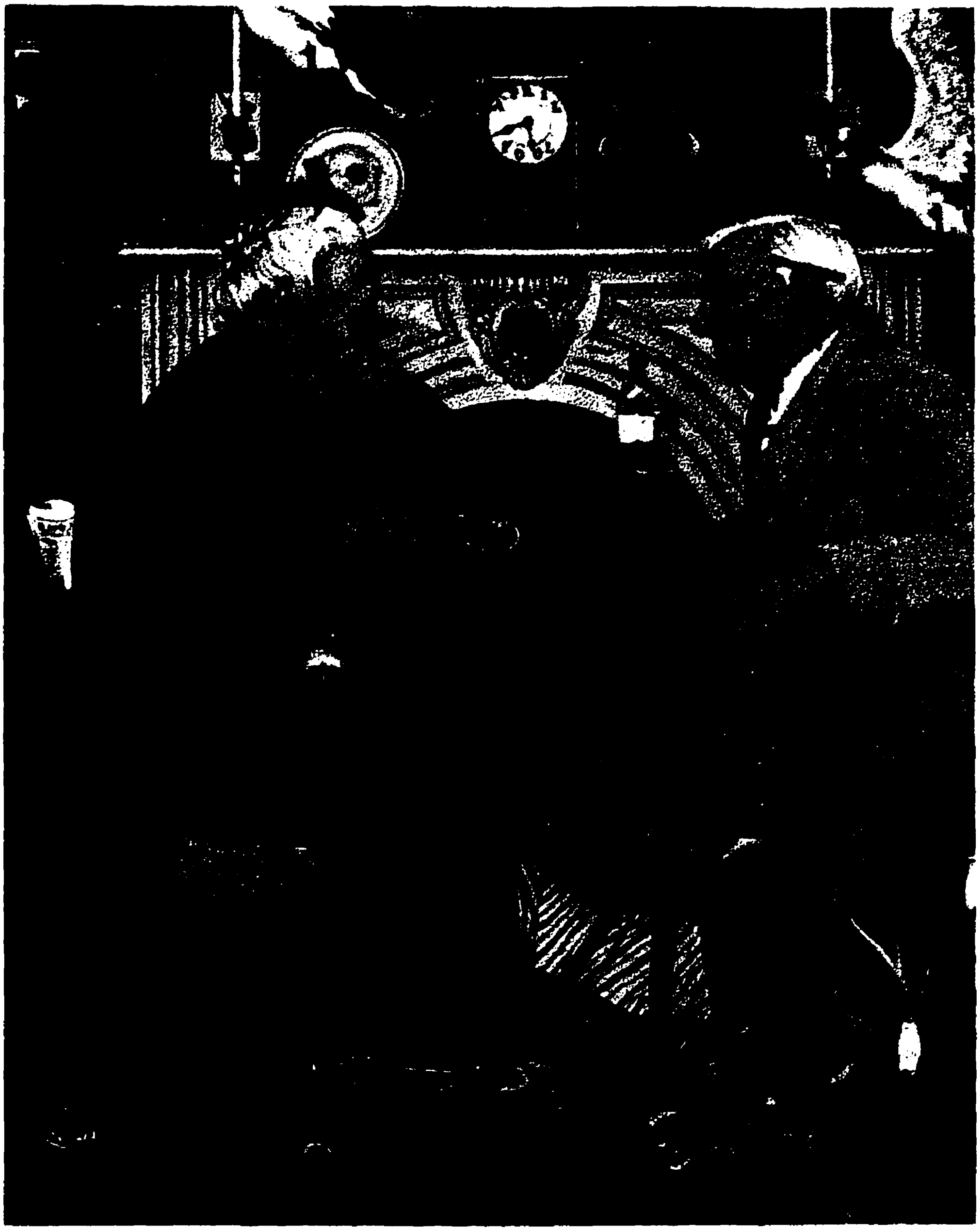
Picture # 6: The Catch



Picture # 7: Traffic Conditions © 1949 The Curtis Publishing Company

Picture # 8: Before the Date

Picture # 9: The Runaway © 1958 The Curtis Publishing Company



Picture # 10: The Game

Appendix 5-ii. Alphabetized List of PNT Stimuli and Practice Items

Stimulus Items

A
ambulance
anchor
apple

B
baby
ball
balloon
banana
basket
bat
beard
bed
bell
belt
bench
binoculars
bone
book
boot
bottle
bowl
bread
bride
bridge
broom
bus
butterfly

C
cake
calendar
camel
camera
can
candle
cane
cannon
carrot
cat
celery
chair
cheerleaders
chimney

C (cont'd)
church
clock
closet
clown
comb
corn
cow
cowboy
cross
crown
crutches

D
desk
dice
dinosaur
dog
door
dragon
drum
duck

E
ear
elephant
eskimo
eye

F
fan
fireman
fireplace
fish
flashlight
flower
foot
football
fork
frog

G
garage
ghost
glass

G (cont'd)
glove
goat
grapes

H
hair
hammer
hand
harp
hat
heart
helicopter
horse
hose
house

I
iron

K
key
king
kitchen
kite
knife

L
lamp
leaf
letter
lion

M
man
map
microscope
monkey
mountain
mustache

N
nail
necklace
nose

N (cont'd)
nurse

O
octopus
owl

P
pear
pen
pencil
piano
pie
pig
pillow
pineapple
pipe
pirate
plant
pumpkin
pyramid

Q
queen

R
rake
ring
rope
ruler

S
saddle
sailor
sandwich
saw
scale
scarf
scissors
seal
shoe
skis
skull
slippers
snail

S (cont'd)
snake
sock
spider
spoon
squirrel
star
stethoscope
strawberries
suit
sun

T
table
tent
thermometer
toilet
top
towel
tractor
train
tree
turkey
typewriter

V
van
vase
vest
volcano

W
wagon
waterfall
well
whistle
wig
window

Z
zebra
zipper

<u>Practice Items</u>

peas
umbrella
waitress
tank
dress
hamburger
mirror
tiger
guitar
whale

Appendix 5-iii. Phonetic Symbols and Descriptions

NAM	IPA	Manner, Place, Voicing	Examples
Consonants			
p	p	stop, bilabial, - voice	*p*et, ti*p*
b	b	stop, bilabial, + voice	*b*et, ri*b*
t	t	stop, alveolar, - voice	*t*ip, pe*t*
d	d	stop, alveolar, + voice	*d*ip, be*d*
k	k	stop, velar, - voice	*c*ap, ba*ck*
g	g	stop, velar, + voice	*g*ap, ba*g*
	ʔ	stop, glottal, - voice	u*h*-*uh* (neg.)
f	f	fricative, labiodental, - voice	*f*at, lau*gh*
v	v	fricative, labiodental, + voice	*v*at, ha*ve*
T	θ	fricative, interdental, - voice	*th*in, ba*th*
D	ð	fricative, interdental, + voice	*th*en, ba*the*
s	s	fricative, alveolar, - voice	*s*ip, le*ss*
z	z	fricative, alveolar, + voice	*z*ip, bed*s*
S	ʃ	fricative, alveopalatal, - voice	*sh*ip, pu*sh*
Z	ʒ	fricative, alveopalatal, + voice	mea*s*ure, rou*ge*
h	h	fricative, glottal, - voice	*h*eat, a*hhh*
C	tʃ	affricate, alveopalatal, - voice	*ch*urch, ca*tch*
J	dʒ	affricate, alveopalatal, + voice	*judge*, garba*ge*
m	m	nasal, bilabial, + voice	*m*ean, la*mb*
n	n	nasal, alveolar, + voice	*n*ear, wi*n*
G	ŋ	nasal, velar, + voice	si*ng*
l	l	liquid, alveolar, + voice	*l*ive, a*ll*
r	r	liquid, retroflex, + voice	*r*ed, ca*r*
w	w	glide, bilabial, + voice	*w*et
y	j	glide, alveopalatal, + voice	*y*et

NAM	IPA	Manner, Place, Voicing	Examples
N		syllabic /n/	butt*on*
M		syllabic /m/	bott*om*
L		syllabic /l/	bott*le*
Vowels			
i	i	high, front, unrounded	b*ea* t
I	ɪ	high, front-central, unrounded	b*i* t
e	e	mid, front, unrounded	b*ai* t
E	ɛ	mid, front-central, unrounded, closed	b*e* t
@	æ	low, front, unrounded	b*a* t
	a	low, front-central, unrounded	b*a* t (Br.)
x	ə	mid, central, unrounded (unstressed)	*a* bout
	ɜ	low-mid, central, more rounded (pre-r or British r)	Be rt, B*er*t (Br.)
^	ʌ	low-mid, back, unrounded (stressed)	b*u* t
a	ɑ	low, back, unrounded, open	b*ough* t
u	u	high, back, rounded, closed	b*oo* t
U	ʊ	high, back-central, rounded, more open	b*oo* k
o	o	high-mid, back, rounded, closed	b*oa* t
c	ɔ	low-mid, back, rounded, more open	b*ough* t (Br.)
Y	ai	diphthong (low>high), front, unrounded	b*i* te
W	aʊ	diphthong (low>high, front>back, unrounded>rounded)	b*ou* t
O	ɔi	diphthong (low>high, back>front, rounded>unrounded)	b*oy*
X	ɚ	retroflex schwa (schwa + r)	butt*er*
R	ɝ	syllabic /r/, stressed	b*ir* d
|		fronted schwa	

Appendix 5-iv. Sample PNT Score Sheet

PHILADELPHIA NAMING TEST SCORESHEET - 175 ITEM VERSION

PNT#:___

	INITIAL ATTEMPT (I)			FIRST COMPLETE ATTEMPT (C)			FINAL ATTEMPT (F)			MPA or MPA√	√
Subject:	RESPONSE	CODE LEVEL 1	CODE LEVEL 2	RESPONSE	CODE LEVEL 1	CODE LEVEL 2	RESPONSE	CODE LEVEL 1	CODE LEVEL 2		
Date:											
Examiner:											
Transcriber:											
Coder:											

LEVEL 1 KEY: √=correct, TA=target attempt, S=semantic, PP=picture part, P=perseveration*, M=mixed / phonological & semantic, B=blend, D=description, O=other, TI-f=target indeterminate/fragment, NR=no response, f=fragment**

LEVEL 2 KEY: S/W(P***)=sound error/word outcome (perseveration), S/NW(P***)=sound error/nonword outcome (perseveration), S/I=sound error/indeterminate outcome (for fragments), MO=morpheme omission (for compounds)

ITEM	WORD	SYLL	FREQ	C/T	I: RESPONSE	I: CODE LEVEL 1	I: CODE LEVEL 2	C: RESPONSE	C: CODE LEVEL 1	C: CODE LEVEL 2	F: RESPONSE	F: CODE LEVEL 1	F: CODE LEVEL 2	MPA or MPA√	√
1	candle	2	L	T											
2	ghost	1	L	C											
3	dinosaur	3	L	T											
4	tree	1	H	C											
5	pen	1	L	T											
6	scissors	2	L	C											
7	cane	1	L	C											
8	comb	1	L	C											
9	thermometer	4	L	T											
10	well	1	L	C											
11	grapes	1	L	C											

NOTE:
*=scorer may refer back to any response (for perseverated words or blends)
**=fragment can go with TA, S, PP, P, & M - eg: S-f=semantic/fragment
***=scorer should refer back only to the previous response (for sound errors)
MPA=multiple phonological attempts - enter "MPA" in column
MPA√=MPA includes target - enter "√" in column
√=target produced w.o. MPA - enter "√" in column

Appendix 5-v. PNT Scoring Codes

(from Roach et al., 1996)

Level 1	Level 2
Correct (✓)	Sound deviation with word outcome (S/W)
Target attempt (TA)	Sound deviation with non-word outcome (S/NW)
Semantic (S)	Sound deviation with indeterminate outcome (S/I)
Mixed (M)	Morpheme omission (MO)
Other (O)	
Blend (B)	
Picture part (PP)	
Perseveration (P)	
Description (D)	
No response (NR)	

Appendix 5-vi. NR Task: Matched Error- and Control-Targets					
Error-Target	**Control-Target**	**Error-Target**	**Control-Target**	**Error-Target**	**Control-Target**
air	guy	bouttoniere	telephone	coffee	gander
apartment	furniture	bow	dog	coffee	money
apparatus	medicine	boy	bum	comb	curl
appointment	equipment	boy	man	comic	paper
apron	lady	buck	guy	cop	guy
army	jacket	building	window	corner	plumber
army	parent	bureau	apron	couple	picture
artist	apron	bureau	lady	crotch	slip
baby	doctor	bush	desk	cuff	man
baby	lesson	carpet	people	cup	girl
back	light	carry	admire	cup	pail
back	young	case	truck	curtain	lady
ball	place	certificate	secretary	curtain	window
basket	bottle	certificate	secretary	daffodil	apartment
be	fit	certificate	secretary	deck	sit
bed	boy	chair	boy	decorate	telephone
behind	around	chair	clothes	deer	men
beige	brown	chair	home	degree	needle
beige	short	chair	home	degree	picture
beleaguer	entertain	chair	milk	deliver	music
belt	chair	chair	phone	design	bacon
bend	pass	checkers	bottle	desk	chair
bend	try	chess	cane	detail	couple
bicycle	window	chess	girl	detail	reason
big	all	chess	man	detail	window
big	full	chignon	paper	dictate	carry
big	two	chocolate	elbow	diploma	telephone
bird	man	chum	game	display	police
biscuit	nothing	church	girl	dock	man
blade	chess	church	hat	doctor	money
blind	coat	church	home	doctor	needle
blind	jar	church	robe	doctor	table
blind	work	cigar	plunger	dog	road
blind	wrench	cigarette	fisherman	dog	truck
blow	got	civilian	another	dog	truck
blow	hold	class	desk	dove	dog
blue	black	clean	most	drawer	army
boat	day	clock	floor	drawer	people
bobby	little	clock	friend	drawn	smoke
bonnet	husband	closet	mirror	dress	has
book	case	clothes	bed	dress	see
book	man	clothes	chair	dress	spray
book	men	clothes	man	dress	stand
boot	one	clothes	slip	dress	tree
bottle	couple	clothes	street	dresser	lady
bottom	little	club	shirt	drink	told
bottom	soldier	coat	hand	duck	game
boutonniere	telephone	coffee	doctor	duck	girl

Appendix 5-vi. NR Task: Matched Error- and Control-Targets					
Error-Target	**Control-Target**	**Error-Target**	**Control-Target**	**Error-Target**	**Control-Target**
duck	skate	handkerchief	coffee	look	give
elderly	musical	handkerchief	spaghetti	look	like
embroider	humorous	hang	got	look	park
examine	lower	hang	look	mad	two
executive	uniforms	hat	boy	mail	dog
eye	deal	hat	shot	main	some
faucet	people	head	have	make	floor
feet	wrench	helmet	apron	man	chair
female	lawyer	her	me	man	cow
feminine	hosiery	here	like	man	desk
figure	carry	his	he	man	guy
fin	man	hobo	jacket	man	rope
fire	van	hold	sit	meat	girl
fish	geese	horn	wrench	medical	overalls
fisherman	balcony	hypodermic	secretary	medical	pyjamas
fisherman	radio	hysteria	ornament	medicine	pyjama
fit	come	ice	leaf	mermaid	apron
fix	comb	ice	stair	mermaid	apron
fix	tell	in	on	mermaid	rubber
flirt	read	in	with	mermaid	water
floor	chair	individual	telephone	military	pretty
floor	man	inject	doctor	mirror	curtain
flower	checkers	innoculate	certificate	mirror	nylon
fly	come	invoice	mirror	model	product
fly	girl	jacket	mitten	morning	lady
fly	wall	jacket	woman	move	live
fool	board	jeans	man	musician	negligee
frock	suit	kerchief	picture	must	want
garbage	lesson	kit	note	nice	old
garden	money	kitchen	basket	nickelodeon	necessary
get	do	lady	picture	nut	way
get	go	lamp	man	old	dead
get	has	lantern	mermaid	old	two
girl	stuff	laurel	chapel	operate	suddenly
give	get	leave	sit	order	police
glass	boss	leave	smoke	overalls	cigarette
glass	man	ledge	man	overalls	telephone
glove	school	let	have	package	scatter
go	come	lieutenant	civilian	pad	shirt
go	did	line	truck	page	man
go	got	liver	police	paint	got
go	want	lobster	lady	pants	boy
grab	hold	lobster	mermaid	pants	maid
guitar	people	lobster	mermaid	pants	man
guitar	people	lobster	tattoo	pants	tool
hair	style	lobster	water	paper	brother
hair	tie	lobster	water	paper	comic
hand	truck	lonesome	little	paper	daughter

Appendix 5-vi. NR Task: Matched Error- and Control-Targets					
Error-Target	**Control-Target**	**Error-Target**	**Control-Target**	**Error-Target**	**Control-Target**
paper	mother	pretty	ticket	sign	side
paper	slippers	pyjamas	cigarette	silly	little
paper	whiskey	pyjamas	coffee	since	just
paraphernalia	handkerchief	pyjamas	peanut	sister	jacket
parlour	people	radio	picture	sit	fold
part	suit	radio	restaurant	sit	see
pass	sit	read	do	skate	hat
peak	half	read	gone	skate	skunk
peck	chair	read	sit	skunk	fire
pecker	pocket	read	think	skunk	man
pedal	mirror	red	pink	sleeve	pants
pencil	lady	rest	talk	slipper	paper
pencil	needle	return	carry	smoke	gun
pencil	woman	revolver	restaurant	smoke	tell
penicillin	certificate	rock	boy	smoke	walk
pennant	army	rock	boy	sneak	sit
pennant	jacket	rock	dog	sock	clothes
perfume	hockey	roller	hockey	soldier	jacket
perfume	lady	roller	woman	some	guy
perfume	mirror	ruler	woman	spaghetti	radio
perfume	paper	run	sit	spaghetti	violin
picture	lady	satchel	mirror	special	woman
picture	people	scale	chair	sports	boy
picture	soldier	scrape	clean	spray	find
picture	stockings	screw	men	spray	stand
picture	wallet	screw	shoe	spray	use
pie	cop	seam	hair	squeegee	window
piece	skunk	secretary	certificate	squirt	shave
pin	maid	secretary	pajamas	squirt	wash
pipe	chair	section	people	stand	boy
pipe	mouth	section	window	stand	read
plant	clock	see	call	stand	take
please	drag	see	go	stand	try
plumber	paper	see	pin	stand	went
plumber	perfume	see	stand	start	go
plumber	perfume	seem	got	stenography	professional
plumber	perfume	serious	different	stool	boy
plumber	window	shelf	stool	stool	man
plunger	mirror	shirt	wrench	stool	pair
plunger	paper	shoe	comb	strap	watch
plunger	person	shoe	girl	string	man
plunger	plumber	shoe	man	string	room
pocket	bottle	shop	smile	sugar	lady
pocket	lady	short	big	suppose	police
pocket	perfume	short	big	surprise	excite
point	want	short	tall	swatter	checkers
police	menu	short	truck	sweater	lobster
police	sugar	should	do	table	basket

Appendix 5-vi. NR Task: Matched Error- and Control-Targets					
Error-Target	Control-Target	Error-Target	Control-Target	Error-Target	Control-Target
table	money	violin	window		
talk	try	visit	open		
tall	good	waiter	police		
tap	bird	waitress	lady		
tap	dog	wall	date		
tee	boy	wall	guy		
tempus	mirror	wall	rug		
thermometer	ingredients	wall	work		
they	he	wash	man		
thing	piece	wash	wink		
thing	tail	washer	saucer		
think	look	waste	job		
think	play	watch	clean		
think	say	way	like		
this	he	way	side		
this	some	wear	blow		
tie	dress	wear	talk		
tie	suit	weigh	read		
tire	man	well	end		
toad	cat	western	different		
toad	man	wharf	man		
today	maybe	what	he		
toiletry	venetian	where	away		
too	else	window	office		
too	much	window	paper		
tool	deer	window	paper		
tool	guy	window	pencil		
tool	man	window	trouble		
torch	face	wink	pull		
transcribe	suppose	wink	tell		
trap	boot	wink	wash		
tray	robe	wire	name		
tree	bill	woman	hammer		
truck	dog	woman	lady		
truck	guy	wonder	enjoy		
truck	man	wood	glass		
truck	road	wood	hand		
try	watch	wrench	bird		
two	back	wrench	man		
two	big	wrench	scale		
under	over	yellow	paper		
uniform	spaghetti	yellow	stocking		
vagabond	pajamas	young	two		
vanity	radio				
verandah	bicycle				
very	over				
vice	not				
violin	radio				