Theoretical and methodological issues in the cognitive neuropsychology of spoken word production

Lyndsey Nickels

Macquarie University, Sydney, Australia

This paper reviews the “state of the art” in the cognitive neuropsychology of spoken word production. It first examines current theoretical issues, specifically addressing the questions: How many levels of processing are there in word production? How does activation flow between levels of processing? What is the relationship between spoken word production and other aspects of language processing? What is the relationship between phonology for lexical and non-lexical forms? The discussion then turns to methodological issues in cognitive neuropsychology, contrasting group, case series, and single case study approaches. Finally, a detailed examination is made of different types of case series approach.

Those clinicians who have met many people with aphasia will appreciate how frequently the overriding concern for these individuals is their ability to speak. No matter how impaired their comprehension, writing, or reading, more often than not it is their (in)ability to verbally express their feelings, wants, and needs that defines the aphasia for them. Similarly, even for those individuals who can communicate effectively using gesture, drawing, and/or writing, their ultimate goal so often seems to be to speak.¹ To (and for) the clinician, the first question then is often “Why?”—What has gone wrong with the processes of speech production for this individual? In order to begin to answer this question, one first needs to understand what processes are involved. Only then can one begin to identify the level at which these processes are impaired for that individual.

¹ I do, of course, appreciate that this is a simplification and moreover that there are individuals for whom this is not true. It may be, for example, that their greatest love was to read, and that without that their life seems devoid of worth, no matter how good their verbal communication skills. In my clinical experience, however, the generalisation holds. I also appreciate that others may feel the primacy of verbal communication to be a reflection of our society, and not an underlying facet of the human psyche. I certainly believe that the impact of a person’s verbal communication impairment on their life, and their perception of this impairment, is of course influenced by the wider social context and attitudes of society as a whole. Nevertheless, the fact remains that many individuals do feel strongly about the importance of verbal communication and these feelings should be respected whatever their origins.

Address correspondence to: Dr L.A. Nickels, Macquarie Centre for Cognitive Science, Macquarie University, Sydney, NSW 2109, Australia. Email: lyndsey@macs.mq.edu.au

Thanks to the authors who have contributed to this volume, who have worked so hard to comply with my demands and provided helpful comments on this paper, and also to the reviewers of their papers. Thanks to Chris Code for inviting me to compile and edit such a special issue, and the production team at Psychology Press for their patience. Max Coltheart also gave helpful comments regarding an earlier version of this paper. This special issue and paper were prepared while the author was supported by an Australian Research Council QEII fellowship.
Unfortunately neither of these steps is straightforward: we neither have agreement on precisely how the proficient speaker produces words, nor do we understand how those processes break down following brain damage. We will examine each of these issues in turn.

We shall be focusing on the production of single words. All of those working in this area acknowledge that isolated production of single words is rare in human communication, and appreciate the importance of sentence-level processing, discourse, and the pragmatics of conversation. It is nevertheless true that at the root of every (verbal) conversation is the retrieval of a single word, and if an individual has an impairment in their ability to retrieve words then their conversation will be affected. Thus, part of the approach to understanding the impact of aphasia on an individual’s life is understanding word retrieval.

THEORETICAL ISSUES IN SPOKEN WORD PRODUCTION

For those involved in research on spoken word production, the publishing of Levelt’s (1989) book *Speaking: From intention to articulation* marked the first serious attempt to specify in detail a complete theory of the process of speech production. In general, psychological theories prior to this date had focused on a relatively constrained central segment of the process (mostly from activation of semantics to retrieval of word form, or thereabouts, e.g., Dell, 1986; Morton, 1970; Stemberger, 1985) or had looked in detail at one section (e.g., phonological encoding, Shattuck-Hufnagel, 1979, 1987). Moreover, there was relatively little “cross-talk” between those focusing on “normal” language processing and those working with aphasia.

Since that time there has been a marked shift, with more emphasis on the use of theories derived primarily from “normal” language processing in the study of impaired language (e.g., Butterworth, 1992; Dell et al., 1997; Nickels, 1997). In addition, studies of impaired language production are increasingly being used to evaluate theories (e.g., Badecker, Miozzo, & Zanuttini, 1995; Dell et al., 1997). This special issue brings together a collection of papers that exemplify this approach—each evaluates theories in the light of data from people with aphasia, and vice versa. In this paper, I will briefly outline these theories and areas of debate (for further discussion see, for example, Hillis, 2001; Nickels, 2001; Nickels & Howard, 2000).

How many levels of processing are there in word production?

Theories of spoken word production have long made the distinction between different forms of information associated with a lexical term: semantic, phonological, syntactic, and morphological. Although there is a general consensus that all of this information is available for a word, there remains a great deal of debate regarding the retrieval of this

---

2 This is not to say that the extent of the effect on conversation is predictable from the retrieval impairment, nor that conversation (and/or communication) will be impaired, merely that if there is a word retrieval impairment, conversation for that individual will differ from conversation before the impairment (e.g., more reliance on gesture, or circumlocution, or use of a higher-frequency, more concrete vocabulary).

3 In other domains, such as reading, theoretical development has always been heavily influenced by cognitive neuropsychological data. Unfortunately, this hasn’t always been the case for spoken word production, with the two strands of research developing relatively independently. Indeed, Levelt and colleagues still appear reluctant to appreciate fully the value of integrating the two approaches (Levelt, Roelofs, & Meyer, 1999b, p. 68).
stored information. The question is whether it all becomes available simultaneously or some types of information are accessed before others.

All theories begin with prelinguistic concepts being used to access the lexical form. In other words, if you wish to convey to someone what the thing was that just bit you, this pre-linguistic concept would activate the semantic, phonological, syntactic, and morphological information associated with “snake” (or “dog”, “spider”, or “child” depending on the circumstances!). Most current theories agree that lexical-semantic and phonological information are accessed separately and that semantic information is accessible prior to phonological information (Figure 1c, e.g., Butterworth, 1992; Caramazza, 1997; Dell et al., 1997; Levelt, Roelofs, & Meyer, 1999a). However, Morton’s (1970) logogen model which was influential in cognitive psychology and cognitive neuropsychology (e.g., Patterson & Shewell, 1987) did not make such a distinction (Figure 1a); activation of phonological output “logogens” by conceptual representations resulted in retrieval of the phonological form (i.e., there was no explicit separate access of lexical-semantic information). Lambon Ralph et al. (this issue) examine this issue using case series methodology and a statistical technique—principal component analysis—to identify whether a theory with just two primary levels (as shown in Figure 1a) could account for the patterns of impairments identified. Only 1 of the 21 aphasic individuals studied by Lambon Ralph et al. showed a pattern significantly different from that predicted by the statistical model (no greater than chance), hence Lambon Ralph et al. argue that this two (semantic and phonological) system theory is all that is required to account for these data. Although this may be true, Lambon Ralph et al. themselves conclude that “almost all models of speech production can explain the data presented here” (p. 80) and it is certain that other authors would disagree that this theory can adequately account for all the current data regarding speech production (see for example the case of PW, Goldrick & Rapp, this issue; Rapp, Benzing, & Caramazza, 1997).

While many current theories (with the exceptions just noted) have broad agreement that there is an “early” lexical-semantic level and a “late” phonological level of processing, how many levels there may be in between, and the nature of those levels, is fiercely contested. Thus Figure 1a–f illustrate six different theoretical stances regarding the levels of processing in speech production, each differing in the number and type of levels of processing involved. Unfortunately it is beyond the scope of this paper to examine each in detail.

In addition to the controversy regarding the number of levels of processing, there is also fierce debate regarding the point at which lexical syntax becomes available (e.g., Caramazza & Miozzo, 1997, 1998; Miozzo & Caramazza, 1997; Roelofs, Meyer, & Levelt, 1998). It is clear that the constraints of sentence structure planning require lexical-syntactic information to be accessed at a relatively early stage of processing. There seems to be a general consensus that lexical-syntactic information regarding

---

4 This is something of a simplification—all theories by necessity must have some mechanism by which you decide what the thing is that you are about to talk about. However, few (if any) theories make it explicit how this mechanism works. In many theories there is a conceptual representation that is distinct from a lexical semantic representation (or lexical concept, Levelt et al., 1999a), but again this is not always made explicit (see Nickels, 2001, for further discussion).

5 While there is also debate regarding morphological representation and access, including whether or not lexical items are represented in a morphologically decomposed form, it is not central to this collection of papers and hence will not be discussed further.
grammatical class, grammatical gender, and the like is retrieved independently from and (usually) prior to (but not necessarily as a prerequisite for) phonological processing being complete (see Nickels, 2001, for a review of this debate). Cognitive neuropsychology has played a central role in this debate, information from individuals with word retrieval impairments (e.g., Badecker et al., 1995; Vigliocco, Vinson, Martin, & Garrett, 1999) providing converging evidence to that from studies of the tip of the tongue state in non-aaphasic subjects (e.g., Caramazza & Miozzo, 1997; Vigliocco, Antonini, & Garrett, 1997). In addition, there have been a number of investigations concerning the influence of grammatical class of words on retrieval in aphasia (although the impact of these studies on theoretical accounts, such as that of Levelt et al., 1999a, is more restricted). For example, in a series of reports Hillis and Caramazza (1995a,b; Caramazza & Hillis, 1991) have reported individuals who have a selective impairment in retrieving either nouns or verbs, and strikingly this impairment is restricted to only one modality (e.g., spoken naming but not written naming). They suggest that these cases provide evidence that stored phonological forms are organised by grammatical class. In this issue, Berndt, Burton, Haendiges, and Mitchum also examine the effect of word class on aphasic word retrieval in different elicitation contexts (e.g., naming to definition vs sentence completion). While they found no reliable effects of elicitation context on performance, grammatical class was once again found to be a significant factor (verbs were harder to retrieve than nouns). Some authors (e.g., Bird, Howard, & Franklin, 2000) have suggested that these effects can be attributed to the fact that verbs and nouns differ not just in their syntactic properties but also in their semantic properties, including differences in imageability/concreteness. Imageability has been shown to be a factor affecting aphasic

Figure 1 (opposite). Diagrammatic representations of five different theories of spoken word production, and their probable relationship to comprehension of spoken and written words and written word production. For each diagram an example (or examples) have been given of authors who have used this architecture.

In this diagram the same terminology has been used for each theory for ease of comparison—in many cases this means that the labels for levels of processing have been changed from those used by the authors. Clearly, the precise properties attributed to each level of processing may not be identical across different authors’ theories even when they have been given the same label here. Readers should refer to the original papers to clarify precise processing and storage assumptions.

No commitment has been made here to the nature of the flow of activation between levels of processing—while each theory takes a stance on this, a priori there is no necessary correspondence between levels of processing and flow of activation.

The majority of theories have not implemented or considered the level of ‘‘prelinguistic concepts’’.

In each case the probable level at which information from other modalities (spoken and written input and written output) converges is indicated with a triple border to a level of processing (with the exception of 1c). Note, however, that not all theories which assume each architecture (in terms of levels of processing) are committed to (nor specify) the level for convergence indicated.

Examples of theories with the levels of processing illustrated:

(1a) Morton’s Logogen model (Morton, 1970, 1985); Patterson & Shewell (1987)
(1b) Ellis & Young (1988)
(1c) Butterworth’s Semantic Lexicon (Butterworth, 1992); Howard & Franklin (1988); Levelt (1989); Nickels (1997).
(1d) Dell et al. (1997); it is unclear for this theory at what level output orthography and phonology might converge. It is possible that the lemma level is modality-independent and orthographic nodes are activated from here, however, it is also possible that this level is modality-specific. The phoneme level is shared for both input and output phonology—hence the double border.
(1e) Caramazza (1997)
(1f) Levelt et al. (1999a)
word production (e.g., Nickels & Howard, 1995b) and are often rated as less imageable than nouns. Bird et al. demonstrate that many supposed grammatical class effects disappear once imageability is controlled, and argue that there is no evidence against the hypothesis that apparent form class effects are simply consequences of semantic differences. Berndt et al. (this issue) reject this as an account of their data as imageability was not found to be a predictor of word production success—unlike grammatical class. However, as they note “the nature of the underlying deficits that give rise to these grammatical class effects remain elusive” (p. 104). It is possible that semantic properties of the stimuli, other than imageability, may still account for the differences between verbs and nouns. For example, Bird et al. note that verbs tend to be loaded on functional and action features, whereas nouns have more sensory properties—hence the association between apparent verb deficits and artefacts deficits.

How are retrieved phonological forms prepared for production?

All of the theoretical accounts sketched in Figure 1a–f are compatible with further intervening phonological assembly processes subsequent to activation of the phonological form and prior to word production—these processes are often referred to as “phonological encoding”. They most commonly involve a “slot and filler” mechanism (Shattuck-Hufnagel, 1979). For example, in Dell (1986), the speech production network comprises two parts, a lexical network (with the general architecture of Figure 1d) where retrieval of the phonological form is represented by a series of activated (position-specific) phonemes, and a wordshape network, where the structure of each syllable is represented (by a series of wordshape header nodes, e.g., CV, CVC). Activated header nodes in the wordshape network select the most active phoneme node for each syllabic position (onset, nucleus, and coda). Levelt et al. (1999a) also make this distinction between representation of phonemes and that of metrical structure (syllabic structure and lexical stress), with their subsequent association in phonological encoding: stressing the need for this process in order for resyllabification to occur for connected speech.6

Some authors suggest that additional processes of phonological encoding are not required. Wilshire (this issue) addresses this subject. Specifically, she examines whether there is evidence necessitating a “dual origin account” for the production of phonological errors (using phonemic misorderings in the phonologically related nonword responses of 22 individuals). She concludes that two levels of phonological processing (retrieval of phonological form and phonological encoding) are not required but that a single level of phonological processing is sufficient to account for the patterns of phonological errors that occur.

Howard and Smith (this issue) also examine the processes involved in phonological retrieval and encoding. In a series of experiments that extend the work reported by Nickels and Howard (1999), Howard and Smith investigate the influence of lexical stress on the word production of aphasic individuals with phonological impairments. As was found by Nickels and Howard (1999), aphasic subjects are more accurate at producing

---

6 “Resyllabification” refers to the fact that when we produce words in connected speech, the syllables are not necessarily those of the individual words as stored. For example, the four syllables in the sentence “I like it all”, will be produced as a different four “I–lie–ki–tall”.

words with first syllable stress (e.g., CAnnon) than words with second syllable stress (e.g., caNOE). They argue that the effect of lexical stress can be clearly localised to the processes of phonological encoding subsequent to the retrieval of lexical-phonological representations (as the effects occur with both words and nonwords). Following Levelt et al. (1999a; Roelofs, 1997), they suggest that metrical structure is only stored for a subset of stimuli—those where lexical stress does not fall on the first syllable. When position of lexical stress is stored rather than assigned by default, the computational effort involved in spelling out the metrical frame results in errors.

How does activation flow between levels of processing?

Not only is there controversy regarding how many levels of processing there are in spoken word production, but there is also no consensus regarding the nature of the connectivity between levels. Thus authors who agree on the levels of processing may differ on how activation flows between these levels. There are two issues here, the continuity of the flow of activation and the direction of flow. The two major competing theories (Dell et al., 1997, Levelt et al., 1999a) differ both in continuity and direction. Dell et al. (1997) describe a theory (which has been implemented as a computer program) that incorporates the levels of processing shown in Figure 1d (without implementation of the concept level). In this theory, activation flows continuously in both directions, thus an activated lexical-semantic node will activate all of those lemma nodes to which it is connected, which in turn will activate the phoneme nodes to which they are connected AND simultaneously feed back activation to the source lexical-semantic node and all other lexical-semantic nodes to which the activated lemma nodes are connected. In contrast to this cascading, interactive activation theory, Levelt et al. (1999a) argue for discrete, feedforward processing in their theory. That is, processing at one level is complete before activation of a subsequent level occurs, and there is no activation of earlier levels of processing by later levels. Rapp and Goldrick (2000) explored the continuum of interactivity between these two extremes, and suggest that a theory with restricted interaction but cascading activation (Restricted Interaction Account: RIA) can more successfully account for speech production data than either highly discrete or highly interactive accounts. This RIA has interaction only between the phoneme and ‘‘L’’ level (which may be interpreted as either phonological form or lemma level and is the only level of processing between lexical-semantics and phonemes). In this issue, Goldrick and Rapp review the challenges for the two ‘‘extreme’’ accounts (e.g., the occurrence of mixed semantically and phonologically related errors and lexical bias) and demonstrate how the RIA can successfully overcome these challenges. They then outline additional challenges for the RIA and argue that these can also be successfully overcome. Other authors have argued for accounts that restrict interaction to between lexical representations and subsequent phonological processes or buffers (see for example Best, 1996; Ellis & Young, 1988, fig 8.3; Nickels 1997, fig 3.4). Best, Herbert, Hickin, Osborne, and Howard (this issue) present evidence from orthographic cueing of aphasic word production that also supports interaction at this level.

---

7 This is a slight modification to Roelofs’ (1997) suggestion required by Howard and Smith’s experimental results. Roelofs (1997) states that metrical structure is only stored when stress does not fall on the first ‘‘stressable syllable’’. See Howard and Smith (this issue) for a detailed discussion of this point.
What is the relationship between spoken word production and other aspects of language processing?

Even restricting ourselves to the level of single words, spoken word production is obviously only one element of language processing. We can also understand and repeat spoken words and repeat strings of phonemes that we have never heard before (nonwords); we can understand and read aloud words and read aloud novel letter strings; we can spontaneously write words, write words and nonwords that are spoken to us, and copy words and nonwords. The question is—how do all of these skills fit together within a theory? The unfortunate fact is that the theories developed to account for each of these skills have all too often developed in isolation (but see Warren & Morton, 1982). Thus, Levelt et al. (1999a) say very little about how spoken word production relates to written word production. Similarly those developing theories of spoken word comprehension, or reading aloud, rarely integrate these theories. In contrast, in the study of the individual with aphasia, the “complete” picture is vital, and therefore has received more attention (e.g., Ellis & Young, 1988; Morton & Patterson, 1980). Consider for example MK, a man with aphasia studied by Howard and Franklin (1988), who when given a picture of a comb to name said /k’omb/ (as opposed to /k’o:mb/). It is only by incorporating writing and reading into a theory of spoken word production that this behaviour can be understood—unable to say the name of the picture, but able to retrieve the written form, MK could then visualise the written form and read it aloud. However, as MK relied on the use of sublexical grapheme–phoneme correspondences to read aloud, this led to the incorrect pronunciation of “comb” as /k’omb/ (as “comb” does not follow spelling–sound correspondence rules).

What then can we say about how these different modalities of language processing relate to each other? Few investigators would propose that each process is entirely separate, and indeed most would concede that all linguistic stimuli converge on the same conceptual representations. However, some authors argue for further modality-independent representations (e.g., lemmas, Levelt et al., 1999a). In addition there are different views regarding how far input and output phonology share representations (and processes), and likewise for orthography. In Figure 1, I sketched six different architectures that have been proposed for spoken word production. These architectures themselves do not require a theoretical stance regarding the relationship to other modalities. However, for at least some of the theories that assume these architectures, the relationship has been made explicit. Hence, Figure 1a–f also shows the level at which heard words, seen words, and written and spoken word production converge, made clear (where possible).

Although all of these theories agree that there are independent stores of orthographic and phonological forms, there is far less consensus about whether each modality needs separate input and output phonological and orthographic representations. Nickels and Howard (1995a) examined the relationship between input and output phonology. They investigated the performance of 15 aphasic individuals on tasks tapping input phonological processing (e.g., minimal pairs, lexical decision) and measures of output phonological processing (e.g., number of phonological errors in word production), and

---

8 Although some authors (e.g., Beauvois, 1982) have argued for multiple semantic systems for different input modalities, these generally refer to tactile, visual (pictures/objects), olfactory, and linguistic. While the “linguistic” modality is usually referred to as the “verbal” modality, verbal semantics can generally be interpreted as lexical-semantics. Few, if any, authors would propose separate semantic information for spoken and written material, nor would they propose separate semantic stores for spoken comprehension and spoken word production.
concluded that there was no relationship between performance in the two modalities. In contrast, Martin and Saffran (this issue) did find a correlation between performance on input and output phonological measures for their aphasic subjects. In a detailed review of the area, they discuss some of the difficulties in comparing performance across input and output modalities, and the predictions of architectures with different functional relationships between input and output phonology. They conclude that the processes subserving input and output phonology are at least functionally linked (and possibly a single system). However, Hillis (2001) concludes her review of this topic by stating that there is evidence (from dissociations) for either ‘‘separate phonological lexicons for input and output, or functionally separate access to phonological representations for comprehending versus speaking’’ (p.195). It seems that there is a long way to go before consensus is reached! Nevertheless, most authors would agree that even if there are functionally distinct input and output representations, at the very least these must be anatomically very close to account for the striking associations in many cases of aphasia.

What is the relationship between phonology for lexical and phonology for non-lexical forms?

As was noted earlier, speakers can easily repeat novel strings of phonemes, which by definition cannot have stored lexical representations. Yet current theories of speech production often fail to make explicit the relationship between this skill and that of production of words. It is possible that the two systems are entirely independent—activating separate phoneme nodes and invoking different phonological encoding procedures. A functionally similar alternative is that the input phonology is sufficiently specified that no phonological encoding is required and motor plans can be retrieved directly. In both of these conditions there would be no necessary correspondence between phonological impairments in spoken word production and impairments in nonword repetition. The two would be entirely independent. However, there have been several reports of individuals with phonological impairments (often classified as ‘‘reproduction conduction aphasics’’) where similar qualitative patterns have been observed for spoken word production and nonword repetition (and reading) (e.g., Caplan, Vanier, & Baker, 1986; Pate, Saffran, & Martin, 1987; Shallice, Rumiati, & Zadini, 2000).

Those accounts that have made explicit the relationship between phonological processing for words and phonological processing for nonwords (e.g., Ellis & Young, 1988; Hartley & Houghton, 1996; Howard & Franklin, 1988; Kay, Lesser, & Coltheart, 1992; Nickels, 1997; Howard, & Best, 1997; Patterson & Shewell, 1987; Shallice et al., 2000) seem to agree that sublexical and lexical phonology converge at the phoneme level (or phonological output buffer in some accounts). Levelt et al. (1999a) also explicitly claim that phonological segments that are activated at input (in the perceptual network) activate the corresponding segments at output (in the production network). Following convergence, the two types of phonology are hereafter identical and subject to the same phonological encoding processes prior to output. This interpretation is supported by identical patterns being found for word and nonword stimuli, when phonological encoding processes are hypothesised to be impaired. For example, Howard and Smith (this issue) found that repetition of stimuli with unstressed initial syllables was more error-prone regardless of whether the stimulus was a word or a nonword. Similarly, all

---

9 Hillis and Caramazza (1991, 1995a) do suggest that sublexical information (from orthography) can directly activate the phonological form, in addition to convergence at the level of the phonological output buffer—see Best et al. (this issue) for further discussion of activation of lexical forms by sublexical information.
those individuals in Nickels (1995) who showed effects of word length on the production of phonological errors in picture naming also showed effects of stimulus length on nonword repetition (Nickels, 1992). 

Best et al. (this issue) discuss this issue in relation to the effects of phonemic and orthographic cueing on aphasic word production. They provided non-lexical cues which were either the first two phonemes of the word (CV cues, e.g., /dp/ for dog), the rime (e.g., /pg/ for dog), or the first two letters of the word (e.g., DO for dog). The provision of these non-lexical cues facilitated picture naming immediately after provision of the cue and this benefit was still evident more than 10 minutes later. How might these non-lexical auditory cues facilitate (prime) word production? One possibility is that the phonemes in the cue activate all the input phonological forms (input morphemes) containing those phonemes and in turn these input forms activate the corresponding output phonological forms (see Best et al.’s fig. 3). The higher level of activation of the output forms results in a greater chance of success in naming. As this priming operates via lexical forms, clearly this is not a possible mechanism for nonword repetition. Nevertheless it remains a plausible mechanism for cueing, as the effectiveness of CV cues was not related to nonword repetition ability for the aphasic participants in the Best et al. study. A second possibility, for explaining effects of non-lexical cues, is that active input phoneme nodes directly activate output phoneme nodes, and similarly active letter/grapheme nodes can activate the phoneme nodes corresponding to their pronunciation.

Thus, where it is made explicit, lexical and sublexical phonology converge at a level where individual phonemes are represented (possibly in a buffer) prior to the phonological encoding of these stimuli. While the available evidence seems to support this view, it would also appear that further research is required (particularly with non-impaired subjects).

**METHODOLOGICAL ISSUES**

Cognitive neuropsychology has been grounded on the study of the single case. In 1988 a special issue of the field’s namesake journal, *Cognitive Neuropsychology*, was devoted to “Methodological problems in cognitive neuropsychology”. Caramazza and McCloskey (1988; McCloskey & Caramazza, 1988) left readers in no doubt that group studies were an unacceptable methodology: “It is argued that only single patient studies allow valid inferences about normal cognitive processes from the analysis of acquired cognitive disorders” (abstract, p.517). Indeed, the majority of the papers published in *Cognitive Neuropsychology* have been single case studies. It might then come as something of a

---

10 All subjects showed significant effects of word length on performance, with the exception of two subjects who were at floor—MK, who was unable to repeat a single nonword; and LAC who was only able to repeat one nonword correctly (of 65)—however, this nonword was monosyllabic.
11 They also included a repetition condition where the whole word was provided.
12 Levelt et al. (1999a,b) do not consider orthography, and therefore quite how phonology might be computed from orthography is an open question. For a review of this issue, see Coltheart et al. (2001).
13 While Caramazza and McCloskey (1988) argue that there is no place for studies of groups of patients within cognitive neuropsychology , Coltheart (2001) is less dismissive. He suggests that “when one is dealing with a domain of cognition about which nothing cognitive neuropsychological is currently known, it is likely to be profitable to begin by seeking to define syndromes within that domain, … Such ideas can be used to develop a modular model of the functional architecture of the system responsible for performance in this cognitive domain.” (p.18). However, he goes on to say “From that point on, the job is to use data from individual patients, not groups of patients, to test that model and any other models which are also formulated.” (p.18). I would argue that for spoken word production “that point” was reached long ago.
surprise, as it did to me as editor, when this special issue on cognitive neuropsychological approaches to spoken word production fails to include a single paper that is devoted to the study of one individual’s word production impairment! How then can we reconcile the two facts—that the studies here do represent the cognitive neuropsychological approach, but yet they all include data from several individuals?

Group studies

The ‘‘group study’’ approach to spoken word production involves selection of a group of individuals on the basis of a defining characteristic (often syndrome category, such as ‘‘Broca’s’’, ‘‘Wernicke’s’’, or ‘‘Conduction’’ type aphasia) and examining qualitative and quantitative patterns of data within and between the groups. Such comparisons are made on the basis of averaged or pooled data across the individuals within a group. Conclusions are then made on the basis of these data of the type ‘‘Broca’s aphasics produce errors of type A OR are worse at task C than task D OR are affected by variable E but not variable F, indicating impairment of type X’’ and likewise ‘‘in contrast Wernicke’s aphasics, produce errors of type B (etc) indicating impairment of type Y’’. However, when data are averaged or pooled across subjects in this way, there is no guarantee that these facts are true for all the individuals within the group (or indeed for any individual within the group). Although it is possible that these conclusions are valid and every individual within that syndrome category will show that pattern, without considering the patterns shown by the individuals as well as that of the pooled data from the group it is simply impossible to know. If the individual patterns are considered, this is the case series or multiple single case study approach. In this approach, a group of patients is indeed selected on some criterion that may be a traditional syndrome category or language descriptor (e.g., ‘‘Fluent aphasia’’, Dell et al., 1997). However, the patterns shown by each individual are important, not just the pattern found in the pooled (or averaged) data (e.g., Nickels & Howard, 1995b). Moreover, in many case series, data are never pooled or averaged across individuals (e.g., Nickels & Howard, 1999). Thus, while group studies are generally ‘‘frowned upon’’ as a methodology for cognitive neuropsychological research, even McCluskey and Caramazza (1988) agree that ‘‘addressing some sorts of theoretical questions requires consideration of multiple cases’’ (p.584).

Case series

Case series, may in fact be of several types.

1. Reviews: Case series across time and researchers. In this case, a series of papers have been published, each a single case study of one individual. Over time it becomes apparent that all these individuals share a common (hypothesised) impairment (or a common symptom). The information from these single cases is then collated and combined in a review. For example, Behrmann, Plaut, and Nelson (1998) review the 57 previously published cases who showed the defining symptoms of letter-by-letter reading.

2. Case series selected for their differences. This approach is used by Goldrick and Rapp (this issue) who report a small case series of two individuals (three individuals in the earlier Rapp & Goldrick, 2000). These individuals were selected because they showed
different patterns of performance across tasks (in terms of accuracy and patterns of errors in spoken word production). These different patterns are then used as tests of different computational models of spoken word production. Clearly to be adequate a computational model must (when lesioned) be able to simulate the performance of the different patterns of performance that occur in the aphasic population. McCloskey (2001, p.602) neatly summarises this approach:

Detailed analyses of deficits can be especially powerful when ... results from several individual patients can be brought to bear on the theoretical issues of interest. In this multiple single-patient study method the goal is to interpret results from all the patients in terms of the same assumptions about normal cognitive mechanisms, by specifying for each patient a form of damage that would lead to the particular pattern of impairment observed for that patient. Note that this research strategy is not aimed at finding multiple patients with exactly the same type of deficit. ... The multiple single-patient study method acknowledges and in fact exploits this diversity, requiring that a theory be capable of accounting for each of the various forms of impairment that may result from damage to the cognitive mechanism under investigation.

(3) Case series selected for their similarity. This is the approach that can superficially appear more like the group study approach. Individuals are grouped according to their performance on one task (or hypothesised nature of their impairment) and the performance on a second task investigated. For example, Berndt et al. (this issue) group their participants according to the nature of their sentence processing abilities. This results in two groups (sentence processing impaired & word retrieval impaired) and one unique individual. The noun and verb production abilities of these individuals are then examined in the context of the groupings. Hence what is being studied is the interaction between hypothesised type of impairment and word class effects in word retrieval (nouns vs verbs). The crucial difference between this approach and the group study approach is that while pooled group data are analysed, no claims are made on the basis of these data without also identifying that the performance of the individuals within the groups also supports these claims.

(4) Case series of relatively unselected individuals. Five of the papers in this issue use this methodology (Best et al., Howard & Smith, Lambon Ralph et al., Martin & Saffran; Wilshire). Aphasic individuals are selected with only very general characteristics, such as impaired word retrieval (Best et al., Lambon Ralph et al., Martin & Saffran) or phonological errors produced in word retrieval (Howard & Smith; Wilshire). Attributes of their performance and/or impairment are then investigated in greater detail.

(i) Effects of task or stimulus attributes on performance: In this volume, Best et al. (who examine the effects of cues of various types on naming), Wilshire (who examines factors influencing contextual errors), and Howard and Smith use this approach. Howard and Smith investigate whether (and how) position of lexical stress affects word production for 12 aphasic participants selected only on the basis that they made phonological errors in speech production. Data from each experiment are analysed both as a group (using ANOVA to determine whether there are significant effects of position

---

14 Two other papers in this issue, Best et al., and Wilshire, also use this approach for part of their analysis. For Best et al., it is the groupings of subjects according to the extent of semantic and phonological impairment used to examine the comparative efficacy of the cues for each group. For Wilshire, it is used in Study 1 to examine the incidence of contextual errors.
of stress on performance overall) but then also for each subject individually, allowing for individual variation to be noted.

One of the problems with analysing individual subject data is how to interpret the fact that one subject may show a significant effect whereas another may not. Can one safely conclude that these two subjects do in fact show a different pattern of performance? To answer this question, Howard and Smith use a test of homogeneity (Leach, 1979) that evaluates whether there is any evidence for patterns of performance varying across subjects. Thus, despite the fact that some individuals show significant effects of lexical stress and others do not, the homogeneity test did not reach significance, indicating that statistically, all 12 subjects show effects of the same size. This statistical technique is clearly a valuable tool when using case series methodology (see also Best et al., this issue), and one that other researchers would benefit from using.

(ii) Comparisons across tasks: A second technique is to compare data from a series of individuals across tasks, with the primary aim of looking for significant associations and dissociations across these tasks. Thus, for example Martin and Saffran (this issue), use a case series of 24 aphasic individuals who made errors in picture naming. They examined whether there were significant associations (correlations) between, for example, measures of phonological processing in output (phonologically related nonword errors) and phonological processing in input (e.g., phoneme discrimination). The patterns of associations and dissociations found were then used to argue for a theory where input and output phonology are (at least) functionally linked (cf. Nickels & Howard, 1995a).

This approach is often used in combination with examining effects of task or stimulus attributes on performance (see section 4i). Thus, Best et al. (this issue) also use this approach to investigate the mechanism by which cues were facilitating spoken naming. For example, they found no significant correlation between naming ability and the size of the facilitation effect, nor between nonword repetition accuracy (for initial phoneme) and efficacy of phonemic cueing, but a significant relationship was found between ability to convert gaphemes to phonemes and efficacy of orthographic cueing. Once again, these patterns provided evidence to support a theoretical interpretation of the cueing effect.

(iii) Finally, Lambon Ralph et al. use a rather different approach. They too investigate the factors affecting spoken word production for a relatively unselected series of aphasic individuals. They provide data on the naming, comprehension, repetition, and reading performance of these individuals. A statistical technique—principal component analysis—was then used to identify any factors underlying the patient data. Two factors were found, and proposed to be semantic and phonological (as the semantic and phonological tasks loaded differently on each factor). These two factors were then used to predict naming accuracy, and (as reported earlier) revealed no difference between observed and expected scores for 20 of the 21 individuals. Of course, the reason why this technique is controversial, is whether a failure to predict the performance of one subject reflects (as Lambon Ralph et al. argue) a chance occurrence given the number of observations, or meaningful data and an important exception (e.g., that there is more to spoken word production than can be explained purely with semantic and phonological factors)? Of course, this relates exactly to the question discussed earlier—how does one decide whether a pattern shown is really different? This is a problem for case series (and single case study) methodology generally, and not simply for principal component analysis.

What benefits might we gain from consideration of case series, as opposed to single cases? In the preface to my 1997 book, I noted that authors of single case studies in the
area of spoken word production often proposed a theory (or modifications to a theory) that accounted well for the particular case they had studied. However, it was less usual for authors to determine whether this theory could also account for the performance of other impaired individuals or indeed unimpaired speakers. This led to a proliferation of theories and a lack of clear progress.\(^{15}\) Since that time it has become more common to relate aphasic data to more general (and less idiosyncratic) theories such as those of Levelt et al. (1999a) and Dell et al. (1997). This has the benefit that the same theories are being revised and tested using data both from experimental investigations of unimpaired speakers and from multiple single case studies of aphasic individuals. This convergence will undoubtedly facilitate the development of these theories.

In addition, as was noted earlier, with a single case study it is difficult to interpret the lack of a significant effect. In contrast, within a case series it is possible to determine statistically whether there is any evidence for individuals showing different patterns of behaviour (as in the example of lexical stress from Howard & Smith, given earlier).

However, this is not to say that there might not be possible detrimental effects of a concentration on case series to the exclusion of the in-depth single case study. One of the most dangerous factors is the restriction on space, and hence detail, that is possible for a paper publishing a case series. In a report of a single case, details may be given of both the task (or skill) in question, and also of every other cognitive skill that may impinge on the subject’s performance. However, in a case series, it simply is not possible to provide all this information. Even when it is provided (in the form of appendices) the full implications are unlikely to be apparent. Take the case of MK (detailed earlier), simply providing data on his naming accuracy and predominant error type (semantic) would not allow the reader to appreciate the complexities of the processing involved in his naming responses. Instead the reader must rely on the integrity of the authors to point out when additional factors are being brought to bear which may not be apparent from the details provided and may affect the validity of the analyses (see, e.g., Nickels & Howard, 1994, footnote 5). Of course, the advantage of the review case series is that here each case has previously been analysed in great detail before the cases are combined. However, standing against this is the difficulty of comparing cases where different tasks and stimuli have often been used. For example, Case A showed no effect of word length on picture naming accuracy for set A, and Case B showed an effect of word length on picture naming accuracy for set B. How sure can we really be that this difference in performance reflects different extent of type of impairment across the two individuals? If Case A was tested using set B (rather than set A) would they now show an effect of length (perhaps set B comprises a greater range of word lengths, or lower-frequency words, which make an effect more likely to become apparent)?\(^{16}\)

In sum then, although not one of the papers in this issue uses data from a single subject to make their case, they are nonetheless cognitive-neuropsychological in their methodology. However, what is also clear is that although the majority can be described as taking a “case series” approach, within that approach there are differences in methodology.

---

\(^{15}\) I appreciate that this is somewhat of a caricature of the situation, and that some authors did relate the impaired performance both to theories derived from experiments with subjects with unimpaired language, and to the performance of other aphasic individuals (e.g., Butterworth, 1992).

\(^{16}\) However, even when individual cases have been tested on the same tasks, equivalent scores have been differently interpreted (for discussion in the area of category-specific disorders, see Best, 2000, pp.102–103).
SUMMARY

In the preface to her edited volume on cognitive neuropsychology, Brenda Rapp (2001) describes it as looking through the window provided by neurological damage onto cognition in order to further our understanding of normal cognitive processing and representation. The study of impairments of spoken word production in aphasia is no exception; the papers in this issue show the variety of patterns of breakdown that occur. The careful analysis and interpretation of these patterns provide the basis for fine-grained inferences about the cognitive processes underlying spoken word production.

REFERENCES


