Notice: This material may be protected by copyright law (Title 17 U.S. Code).
Journal Title: Aphasiology

Volume: 10 (2)
Issue: 
Month-Year: 1996
Pages: 202-215

Call #: RC425.A675
Location: PATERNO-4

ILL #: NYCE8344

Ariel: 128.59.88.36
Patron: Froud, Karen
RLG: NYCE

Notes:

From:
Penn State Interlibrary Loan Lending
127 Paterno Library, Curtin Road
University Park, PA 16802

Fax: 814-863-2653
Ariel: 128.118.152.188

12/3/2003

Ship To:
Teachers College Library
Interlibrary Loan - Room 18/Box 18
525 W. 120th Street
New York, NY 10027 USA

Fax: 212 678 3092
The maximum difference is 6 distinctive features (df) plus a vowel change, which makes all three phonemes of the pair different.

<table>
<thead>
<tr>
<th>Cond</th>
<th>Differences</th>
<th>Features changed</th>
<th>No. of errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3df changed; both consonants + vowel</td>
<td>6 + V</td>
<td>/4</td>
</tr>
<tr>
<td>2</td>
<td>2df, both consonants + vowel change</td>
<td>4 + V</td>
<td>/4</td>
</tr>
<tr>
<td>3</td>
<td>1df, both consonants + vowel change</td>
<td>2 + V</td>
<td>/4</td>
</tr>
<tr>
<td>4</td>
<td>3df, both consonants</td>
<td>6</td>
<td>/4</td>
</tr>
<tr>
<td>5</td>
<td>2df, both consonants</td>
<td>4</td>
<td>/4</td>
</tr>
<tr>
<td>6</td>
<td>1df, both consonants</td>
<td>2</td>
<td>/4</td>
</tr>
<tr>
<td>7</td>
<td>3df, one consonant</td>
<td>3</td>
<td>/4</td>
</tr>
<tr>
<td>8</td>
<td>2df, one consonant</td>
<td>2</td>
<td>/4</td>
</tr>
<tr>
<td>9</td>
<td>1df, one consonant</td>
<td>1</td>
<td>/4</td>
</tr>
<tr>
<td>10</td>
<td>Same</td>
<td>0</td>
<td>/12</td>
</tr>
</tbody>
</table>

**Practice items**

<table>
<thead>
<tr>
<th>No.</th>
<th>Items</th>
<th>Response</th>
<th>No.</th>
<th>Items</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>gsg</td>
<td>D s</td>
<td>P6</td>
<td>bbs</td>
<td>D s</td>
</tr>
<tr>
<td>P2</td>
<td>gak</td>
<td>D s</td>
<td>P7</td>
<td>gas</td>
<td>D s</td>
</tr>
<tr>
<td>P3</td>
<td>kk</td>
<td>S d</td>
<td>P8</td>
<td>kik</td>
<td>D s</td>
</tr>
<tr>
<td>P4</td>
<td>sgb</td>
<td>D s</td>
<td>P9</td>
<td>bbs</td>
<td>D s</td>
</tr>
<tr>
<td>P5</td>
<td>trg</td>
<td>S d</td>
<td>P10</td>
<td>arg</td>
<td>D s</td>
</tr>
</tbody>
</table>

**Test items**

<table>
<thead>
<tr>
<th>No.</th>
<th>Items</th>
<th>Response</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ses</td>
<td>bbb</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>tsk</td>
<td>bbb</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>dat</td>
<td>bbb</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>log</td>
<td>S d</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>tdd</td>
<td>bbb</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>dst</td>
<td>bbb</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>kbb</td>
<td>S d</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>bbs</td>
<td>bbb</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>peb</td>
<td>bbb</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>dib</td>
<td>bbb</td>
<td>7</td>
</tr>
<tr>
<td>11</td>
<td>sat</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>12</td>
<td>lmb</td>
<td>bbb</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>pdd</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>14</td>
<td>abg</td>
<td>bbb</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>gog</td>
<td>bbb</td>
<td>8</td>
</tr>
<tr>
<td>16</td>
<td>bbb</td>
<td>S d</td>
<td>10</td>
</tr>
<tr>
<td>17</td>
<td>bbb</td>
<td>bbb</td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>ses</td>
<td>bbb</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>peb</td>
<td>bbb</td>
<td>5</td>
</tr>
<tr>
<td>20</td>
<td>kes</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>21</td>
<td>dek</td>
<td>bbb</td>
<td>3</td>
</tr>
<tr>
<td>22</td>
<td>bbb</td>
<td>bbb</td>
<td>6</td>
</tr>
<tr>
<td>23</td>
<td>pbb</td>
<td>bbb</td>
<td>9</td>
</tr>
<tr>
<td>24</td>
<td>gbb</td>
<td>bbb</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>No.</th>
<th>Items</th>
<th>Response</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>us</td>
<td>bbs</td>
<td>10</td>
</tr>
<tr>
<td>26</td>
<td>pbb</td>
<td>bbs</td>
<td>6</td>
</tr>
<tr>
<td>27</td>
<td>hbb</td>
<td>bbs</td>
<td>9</td>
</tr>
<tr>
<td>28</td>
<td>gas</td>
<td>bbs</td>
<td>4</td>
</tr>
<tr>
<td>29</td>
<td>kbb</td>
<td>bbb</td>
<td>2</td>
</tr>
<tr>
<td>30</td>
<td>sbb</td>
<td>bbb</td>
<td>7</td>
</tr>
<tr>
<td>31</td>
<td>sbb</td>
<td>bbb</td>
<td>5</td>
</tr>
<tr>
<td>32</td>
<td>dbb</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>33</td>
<td>gbb</td>
<td>bbb</td>
<td>1</td>
</tr>
<tr>
<td>34</td>
<td>seb</td>
<td>bbb</td>
<td>5</td>
</tr>
<tr>
<td>35</td>
<td>dss</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>36</td>
<td>kbb</td>
<td>bbb</td>
<td>8</td>
</tr>
<tr>
<td>37</td>
<td>sbb</td>
<td>bbb</td>
<td>9</td>
</tr>
<tr>
<td>38</td>
<td>tbb</td>
<td>bbb</td>
<td>6</td>
</tr>
<tr>
<td>39</td>
<td>kbb</td>
<td>bbb</td>
<td>3</td>
</tr>
<tr>
<td>40</td>
<td>bbb</td>
<td>bbb</td>
<td>5</td>
</tr>
<tr>
<td>41</td>
<td>pbb</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>42</td>
<td>gbb</td>
<td>bbb</td>
<td>2</td>
</tr>
<tr>
<td>43</td>
<td>sbb</td>
<td>bbb</td>
<td>1</td>
</tr>
<tr>
<td>44</td>
<td>ptd</td>
<td>bbb</td>
<td>8</td>
</tr>
<tr>
<td>45</td>
<td>dtp</td>
<td>bbb</td>
<td>10</td>
</tr>
<tr>
<td>46</td>
<td>btp</td>
<td>bbb</td>
<td>7</td>
</tr>
<tr>
<td>47</td>
<td>tss</td>
<td>bss</td>
<td>10</td>
</tr>
<tr>
<td>48</td>
<td>gss</td>
<td>gss</td>
<td>4</td>
</tr>
</tbody>
</table>

Clinical Forum

Psycholinguistic assessments of language processing in aphasia (PALPA): an introduction†

JANICE KAY, RUTH LESSER and MAX COLTHEART
University of Exeter, Exeter, UK

Introduction

PALPA is designed to be a resource for speech and language therapists and cognitive and clinical neuropsychologists who wish to assess language processing skills in people with aphasia. We believe that PALPA can make a substantial contribution to the investigator/therapist’s resources for examining people with aphasia. The comments made by a large number of aphasia therapists throughout the UK, other parts of Europe, and Australia and Canada—some of whom have been using research versions of the battery—have been encouraging. PALPA already seems to have brought a new approach to the clinical examination of individual patients with dysphasia, one which is in tune with the philosophy of considering language assessment as an iterative procedure of hypothesis testing.

At the same time, we must make it clear that it provides materials for only a circumscribed part of what needs to be assessed when one investigates the language abilities of a person with dysphasia. In the first place it is concerned primarily with language as a complex series of mental processing steps and makes a somewhat artificial distinction between this and what we do when we use language to communicate. At present there is a substantial gap between the assessment of language processing as a mental activity and language used as a means of communication in everyday life. But as assessment of the latter through conversational analysis is at an early stage of development (see, for example, Gerber and Gurland 1989, Lesser and Milroy 1993), the investigator/therapist has no alternative but to assume that it is acceptable to examine language processing and language use as separable. Current psycholinguistic investigations of language processing (and PALPA is no exception) may seem to maintain this gap by using some tasks which rarely form part of the customary use of language. One of these is reading single words aloud, an activity generally performed only in restricted situations (by mothers or teachers of young children, for example). Another is

† Reproduced from booklet 1 of 5.
PALPA is available from Erbaum (UK) Taylor & Francis, 27 Church Road, Hove, East Sussex BN3 2FA, UK.
should be tailored to those that are appropriate to the hypothesis under investigation. Once a hypothesis about which modules are dysfunctioning is set up, the clinician can then plan a treatment programme which would be appropriate to restoring, reorganizing or compensating for the impaired processes. PALPA does not specify which treatment programme should be carried out; rather its aim is to provide a firm grounding for an understanding of a particular processing disorder on which any treatment programme must be based.

A theoretical background

To appreciate how PALPA works, and to use it effectively, the model of language processing upon which each assessment is based must be understood, and one of the aims of this introductory manual is to provide a brief (and therefore necessarily basic) understanding. We need to explain what the components of the model are, what the pathways of communication between these modules are, and what language-processing function depends upon each module or pathway. At the end of this paper we have provided a list of further reading that will allow you to explore, in your own time, the complex theoretical issues on which the approach is based.

Recognizing printed words

If someone asks you which of the two letter-strings meach and peach is a word, you can perform this task rapidly and effortlessly. How do you do it? What are the mental processes involved in making such a decision? The letter-string meach is a perfectly well-formed item which could have become a word of English, but never did, unlike peach. Given this, the only way you could be carrying out this task is by searching through all the words of English that you know, and finding that peach is among this large set of words, whereas meach is not. Here you are dealing with one particular sort of knowledge about words, namely orthographic knowledge—knowledge about spelling. It is as if you are consulting a kind of mental dictionary that contains the spellings of all the words you know. Recognizing a word consists of finding it in this system. The system you use to recognize visually presented words is often called an 'orthographic input lexicon': orthographic because it deals with the spellings of words; input because it is for taking in information rather than producing it; and lexicon because it is a list of words, like a dictionary. It differs from the usual kind of dictionary, however, because it contains only one kind of information about words: orthographic information. It does not know anything about meanings or pronunciations. These forms of information are stored in other modules of the language-processing system, which will be discussed below. Figure 1 is our first step towards building up a complete diagram of the whole language-processing system; it is a picture of just one fragment of that system. Every word that the reader knows is represented by a separate entry in this orthographic input lexicon. According to this view, non-words like meach, no matter how similar to words they are, do not have entries in the orthographic input lexicon. As there is good evidence that the letters of a word must first be recognized before the word itself can be recognized, there must be a system for recognizing letters that operates before access to the orthographic input lexicon. This is called 'abstract letter identification' because, to this procedure, an A is an A: it ignores letter font and letter case.

Rationale

PALPA stands for 'psycholinguistic assessments of language processing in aphasia'. It consists of 60 assessments designed to help to diagnose language processing difficulties in individuals with acquired brain damage. As its name suggests, PALPA applies a psycholinguistic approach to the interpretation of processes concerned with the recognition, comprehension and production of spoken and written words and sentences. The approach is based on the assumption that the mind's language system is organized in separate modules of processing, and that these can be impaired selectively by brain damage. PALPA aims to provide information about the integrity of these modules, to find those in which the aphasic person seems to be functioning below normal and those which appear to be continuing to function normally or near-normally. It is important to realize that PALPA is not designed to be given in its entirety to an individual—rather the assessments
This minimal information-processing system cannot do much: it cannot understand words or speak them, as it does not contain information about word meanings or word sound-forms. However, it is not entirely ignorant of language, as it can recognize printed words, and can decide whether a string of letters is a real word or not.

Understanding printed words

Not only do you know that the letter-string *peach* is a word, but you also know that it means a kind of fruit. Thus, you possess not only orthographic but also semantic knowledge about this letter-string. Knowledge of the meanings of words is stored in a module of the language-processing system which we will call the 'semantic system'; this contains information about the meanings of all the words known by the person concerned. To get to this system from a printed word you have to recognize the word first; that is, to find its entry in the orthographic input lexicon. Having done this will enable you then to find its representation in the semantic system. So, as Figure 2 shows, the orthographic input lexicon is a gateway to the semantic system. This figure shows a slightly larger fragment of the language-processing system, a fragment that can not only recognize printed words but can also understand them (though not speak them).

A digression: boxes and arrows

At this point it might be worth digressing briefly to explain what the boxes and arrows in diagrams like Figure 1 and Figure 2 are actually supposed to represent. In all of the box-and-arrow diagrams you will see here, the boxes have two types of function. Firstly, they are repositories of information; for example, the box labelled Semantic System in Figure 2 actually contains semantic information. Secondly, the boxes are also processors of information. It is not enough for a language-processing system to include a component containing semantic information; it will also be necessary for there to be a procedure for finding the desired information within that component. To take another example: we have said that the way someone decides that *peach* is a word is by finding its entry in the orthographic input lexicon. For this to happen, not only must there be a representation of this word in the orthographic input lexicon, but there must also be a procedure which enables this representation to be found among the tens of thousands of other word-representations there. A really precise description of the language-processing system, then, would depict it in a way that distinguished between systems of representations and the procedures that are used to find entries in these systems. However, this degree of precision is something we do not need here, so we simply collapse the two aspects of boxes into one. Any box you see might therefore actually consist of a collection of representations, or a system for processing representations, or both.

What about the arrows? These are to allow communication between the boxes. Once the entry for the word *peach* has been found in the orthographic input lexicon, some form of communication from this system to the semantic system will be needed if the reader is to know what the word means. This communication is represented by the arrow from the orthographic input lexicon to the semantic system in Figure 2. Exactly how these channels of communication might work is something about which little is known. Fortunately, nothing important about the use of PALPA depends on this. However, it might be worth briefly describing a couple of ways in which one could imagine these arrows working.

Suppose that each of the entries in the semantic system had some kind of unique and quite arbitrary code associated with it—a code that one might think of as a unique number, so that *peach*, for example, might be word number 1873 in the semantic system. Now, if the code 1873 is stored along with the word's entry in the orthographic input lexicon, then once the entry for *peach* has been found in that lexicon, the code stored there could be transmitted to the semantic system, and the processor in the semantic system would then find the right meaning by finding meaning No. 1873. Another completely different way of thinking about how the arrows might work is to imagine that there is a direct connection between each entry in the orthographic input lexicon and its corresponding entry in the semantic system. Now an arrow is a large bundle of private lines, not a single channel for transmitting a word's unique code.
of the word as part of an utterance. These may include a ‘buffer’ system of temporary storage in which the string of sounds that will form the word is assembled, allowing their allophones within the context of the word to be specified and neuromuscular programming for the utterance to be carried out. For simplicity’s sake we have subsumed all these operations under the term ‘speech’ in Figure 3, although we expand on why buffers are needed in the later section on buffer storage. It is important to note, however, that disturbances at any of these further stages can interrupt word-realization for speech in aphasia, and that distinguishing the level or levels at which the difficulties lie can be critical for planning therapy for anoma (see Lesser 1987).

We reached this diagram by thinking about how we recognize words, how we understand words, and how we speak them spontaneously. Not surprisingly, then, the information-processing system in Figure 3 can do all of these things. There is, however, something else it can do that it was not specifically constructed to do: it can read aloud. If there is a pathway from orthographic input lexicon to semantic system, and another pathway from semantic system to phonological output lexicon, then the use of both pathways will provide a route from print to speech; that is, a procedure for reading aloud. So this system can not only recognize and understand printed words and speak spontaneously, it can also read aloud.

Reading aloud

Is this all we need to say about reading aloud? Is Figure 3 sufficient to describe everything about reading aloud? No. In the case of non-words like meach, not only can you recognize that these are not words, you can also read them aloud, which the system in Figure 3 cannot do. The non-word meach will not be found in the orthographic input lexicon, so there will be no communication from this lexicon to the higher components of the system, and in particular nothing will get to the phonological output lexicon. Even if it did, that would not help: the phonological output lexicon contains only words, so could not be used to utter a non-word. Hence the system in Figure 3, when confronted with a printed non-word, would remain mute; it could not read non-words aloud. Yet non-brain-damaged (and not necessarily practised) readers can do this. Hence we need to add something to Figure 3 to account for how non-words are read aloud.

What is required is some procedure which will relate the spellings of non-words to their sounds: a set of ‘spelling-to-sound’ rules. For example, if you know the three rules m → /m/, ea → /i/, and ch → /ʃ/, applying these rules to the letter-string meach will give you the right phonological output, /mitʃ/. Hence the developing model needs to be supplemented with a letter-sound rule system, as in Figure 4.

Now the model has two separate procedures for reading aloud: it is a ‘dual-route’ model. One of these procedures goes from the orthographic input lexicon to the phonological output lexicon (via the semantic system), so we will call this the ‘lexical route’. The other does not use lexicons at all, so we will call this the ‘non-lexical route’.

As we have explained, the lexical route cannot read non-words at all (because lexicons contain only words). The non-lexical reading route also has a limitation: as it uses spelling-sound rules, it cannot succeed with words that disobey such rules—the irregular or ‘exception’ words of English. It will read pint as if it
rhymed with mint, it will pronounce the ch in yacht and both is in colonel, because this is what the rules prescribe. These are known as ‘regularization’ errors. On the other hand, if the input is a ‘regular’ word like mint or boat or army, the non-lexical route will respond correctly, as it will when the input is a non-word.

This means that, to read aloud adequately, the language-processing system needs both procedures. Only the lexical procedure can read exception words correctly. Only the non-lexical procedure can read nonwords correctly.

Are two procedures enough to explain what we know about reading? Possibly not. Schwartz et al. (1980) described a dementing patient who had almost no word comprehension left; that is, she could make little or no use of her semantic system. However, she could read aloud rather well, including exception words such as leopard. How did she read exception words aloud so well? Not by using spelling-to-sound rules (which would result in regularizations of exception words); and not by using the lexical route via the semantic system, because she had a very severe semantic impairment. These and similar findings (e.g. Funnell 1983) lead us to introduce a third reading route or, rather, to subdivide the lexical route, by adding a pathway that goes directly from orthographic input lexicon to phonological output lexicon, by-passing semantics; this is shown in Figure 5.

Recognizing objects and pictures

Recognizing a peach, or a picture of a peach, is not purely a linguistic activity. Nevertheless, we need to add this capability to our model of the language-processing system simply because picture-word or object-word matching are commonly used as tests of comprehension in aphasia assessments (including PALPA); so we have to say something about picture and object comprehension.

We are going to assume (though not everyone does) that the semantic system used to understand printed words (and spoken words) is the same as the semantic system used to understand pictures and seen objects. If this is so, then to understand a picture or a seen object the viewer will have to gain access to the representation of that picture or object in the semantic system of Figure 5. To achieve this there would need to be a system capable of recognizing familiar pictures or objects and then communicating with the semantic system—something analogous to the orthographic input lexicon of Figure 1, but operating upon pictures or objects rather than words. We will call this the ‘visual object recognition system’, and see it as containing entries which are structural descriptions of objects—there will be an entry for ‘giraffe’, for instance, that specifies long-neckedness, long-leggedness, and spottedness. Recognition of an object or picture will have occurred when the viewer has succeeded in matching up the visual features of the viewed stimulus with the details of one of the structural descriptions in the visual object recognition system. Once this is done, communication to the
Recognizing and repeating speech

The system in Figure 5 can process print and pictures, but not speech: an obvious lack is a 'phonological input lexicon', so it is time to add this, and a preceding system of 'acoustic analysis' that is analogous to the abstract letter-identification system of Figure 1. There will need to be an arrow from the phonological input lexicon to the semantic system, to allow spoken words to be understood rather than merely recognized, and an arrow from phonological input lexicon to phonological output lexicon, for repetition of words.

We can, however, repeat not only words but also non-words, which could not happen if repetition were accomplished only via a route from an input to an output lexicon. Hence it must be possible to repeat by going directly from acoustic analysis to speech (via a process of 'acoustic-to-phonological conversion'), by-passing lexical systems.

This gives us Figure 7. Just as it has a lexical and a non-lexical route for reading aloud, it has a lexical and a non-lexical route for repeating. Why? We have already seen that our ability to repeat non-words requires a non-lexical repetition route. It is often found in aphasic patients that words can be repeated whilst non-words cannot; this could not happen if the only repetition route were non-lexical, so there must be a lexical repetition route as well.

Spelling and writing

Spelling and writing are part of language-processing, a part that the system in Figure 7 cannot accomplish. Hence it needs further elaboration. We already have a phonological output lexicon for producing spoken words, so we will add an 'orthographic output lexicon' for producing written words. So spontaneous writing consists of communication from semantic system to orthographic output lexicon, and retrieving from that output lexicon the written forms of the appropriate words. This would not explain our ability to write to dictation spoken non-words like meach, so we have to add a system of 'sound-to-spelling rules' too.

The arguments here are exactly those discussed earlier in relation to reading. If spelling depended only on sound-to-spelling rules, we would spell exception words incorrectly, perhaps spelling *yacht* as *yot*. If spelling depended only on an orthographic output lexicon, we could not spell non-words to dictation; so both spelling routes must be included in the model, and this gives us Figure 8.

Buffer storage

Suppose someone asked you to write down the word *rhinoceros*. Its spelling is one of the entries in the orthographic output lexicon, so you retrieve this spelling and begin to write the word. Now, it is going to take you several seconds to get all ten of these letters down on paper, so the spelling specification you have retrieved from the orthographic output lexicon will need to remain available for several seconds. Where? It is permanently available in the orthographic output lexicon, but you do not want to have to keep going back to that system and searching through the tens of thousands of words in it every time you want to know what the next letter in the word is. It would be a great help if you had a means of storage that could temporarily hold just the letters of the one word you want to write until you have
finished writing it. Systems for the brief temporary storage of information that is currently being worked on are called 'buffers', and studies of normal spelling and of acquired dysgraphia suggest that there exists an 'orthographic output buffer' that works in just this way.

Exactly the same argument can be made about spoken rather than written output—saying *rhinoceros* is no more an instantaneous event than writing it—so this leads to the concept of a 'phonological output buffer'. We need a 'phonological input buffer' too. A printed word is permanently available to processing mechanisms until they have completed their job, but speech is fleeting—so a spoken word must be preserved in a temporary memory for long enough that the task of auditory word recognition can be completed by the phonological processing systems. Thus, we arrive at Figure 9, the model of language-processing in its final form.

Is this really how single words, non-words and pictures are processed?

We have already indicated ways in which the model set out in Figure 9 is not universally accepted, and there are yet other points of controversy: some people would want an arrow from visual object recognition to phonological output.
Although it may seem odd to say so, this is not crucially important for present purposes. In combination with the PALPA materials this model can be used to interpret various patterns of language impairment, to guide the choice of assessments, and to assist in the design of treatment programmes. We are sure that the model is extremely helpful for all three of these endeavours, even if it turns out that there are various ways in which it is wrong. To try to understand why a particular patient can succeed in certain language-processing tasks while failing in others by relating the pattern of successes and failures to the Figure 9 model forces one to think hard, and with precision, about the patient’s performance, and the model tells one how to do so. Data gathered by the PALPA assessments suggested by the model are going to be informative simply because basic distinctions upon which the model is based (such as word vs. non-word or regular word vs. exception word) are known to have major influences upon patient performance and so must be basic to any model that aspires to explain normal and disordered language-processing.

Sentence processing

Aphasic patients may be able to recognize and to understand single written and spoken words, yet have difficulty in understanding the meanings of whole sentences. PALPA assesses this by using sentence–picture matching tasks. One of the features we have included in designing these tasks is whether the sentence is plausibly semantically reversible; that is whether, if one interchanged the two noun phrases, the resulting sentence would still be pragmatically acceptable. This is essential; if the sentences are not pragmatically reversible, a patient with a sentence-processing impairment could work out what a sentence must mean by working out the only semantically and pragmatically possible combination of the sentence’s words. For example, suppose a patient has lost the ability to identify the subject and the object of a sentence. In the non-reversible sentence, ‘The man likes the steak’, as steaks cannot like anything, it must be the man that is liking and the steak that is liked, and the patient could work this out even if unable to assign syntactic categories to the words in the sentence. But in the reversible sentence, ‘The man likes the woman’, this non-syntactic strategy will not be useful, and so performance will be poor on sentence–picture matching tasks if the patient has a sentence comprehension impairment.

If a patient does poorly on sentence–picture matching, two possible reasons for this can be explored. The first possibility is that sentences are not understood because single words are not successfully understood. Test [57] helps to evaluate this possibility, as it provides some information about how well the subject understands individual verbs and adjectives that are used in the sentence–picture matching tasks [55, 56]. Lexical semantic influences are controlled in the sentence–picture matching task, partly through using a limited set of referents (all animate), and partly through incorporating the dimension of directionality of motion in some of the verbs used (e.g., following, leading). Of course the results of picture and word semantics tests should also be taken into consideration here, as these tasks measure single-word comprehension.

Other factors are also controlled for in our sentence–picture matching task: these are whether the predicate is a verb or an adjective, active-passive mood, length of sentence, and sentence complexity in terms of whether all the elements are made
specific in the spoken or written sentence, or whether the person listening or reading the sentence has to fill in 'gaps'. For example, in the sentence 'The man is demonstrating what to do', there is a presumed gap in that the person who will be doing what has been demonstrated is not grammatically specified (the same applies even when the doer is the same as the subject of the sentence, as in 'The man is asking what to do'). This relates to a linguistic theory which involves what are called 'PRO structures at a d-structure level', or 'traces at an s-structure level' (see Caplan and Hildebrandt 1988 for a discussion of this as applied to syntactic comprehension in aphasia). The point of including these distinctions in PALPA is to see whether some patients have more difficulty with these more complex types of sentences than with others. Moreover, as well as comparing verbs and adjectives as predicates (given that it has been suggested that some patients have specific difficulties with verbs), we have included a further picture-choice task in which the predicates are locative prepositions (58, 59); this task also allows for the influence of degree of animacy/abstractness to be examined.

A second possible reason for poor performance on sentence—picture matching is impaired memory. It is often argued that sentences need to be held in working memory while syntactic processing is carried out; this kind of memory seems to be different from the memory needed to store a string of words (such as a phone number) which are not syntactically related, and which do not require a hierarchical 'tree structure' to be formed. If patients have a poor working memory for sentence processing they will not be able to understand sentences precisely, although they may be able to extract some gist in the form of a string of referents. One of our tasks investigates reduced memory span in sentence processing by giving the subject a sequence of words which could take the roles of noun phrase and verb phrase in an anomalous sentence. Instead of having to understand the sequence, the patient simply needs to hold it in memory, and this ability is assessed by requiring the patient to point, in the correct order, to the pictures that correspond to each of the words. If the patient can do this with word sequences of a particular length, but fails in sentence—picture matching with sentences of that length, the problem with sentence processing cannot be ascribed to a failure of memory. Of course, the results of some of our auditory processing tasks (77 to 133 inclusive) should also be taken into consideration here, as these tasks also measure memory for linguistic stimuli.

As well as compiling our own sentence-processing tasks, we also gave Bishop's (1982) Test for reception of grammar to a selection of our non-brain-damaged subjects. This test examines performance on a variety of different sentence constructs and is designed for assessing receptive grammatical abilities in children with a variety of specific and non-specific language disorders. Bishop indicates, in the manual that accompanies TROG, that centiles for 12-year-olds can be used for adults passing up to 18 blocks. When the TROG sentences were presented auditorily to 17 of our non-brain-damaged subjects, their mean score was 18.5 blocks passed (standard deviation = 107), which falls roughly between the 50th and 60th centile. The results are broadly similar for sentences presented in written form: 15 subjects passed 18.5 blocks (standard deviation = 1.46).

Selection of assessment tasks

Each of the 60 tasks that comprise the PALPA battery is designed to help to illuminate the workings of specific components of the language-processing model described here. The tasks are well founded and, in the main, are based on those described in the neuropsychological and experimental literature and do not introduce new theoretical constructs or experimental techniques. Even with 60 tasks, there are bound to be tests of language processing that we have not included. We have omitted any assessment of sentence production, for example, because in this case we felt that there were already readily accessible analyses of this language function (e.g. Saffran et al. 1989). We have also not chosen to assess many important aspects of language use, such as inference-making. Just as the theoretical basis of the battery assumes a modular organization, the organization of the battery can be seen as modular, and can be added to as further assessments of language use are developed. Neither have we chosen to look at 'on-line' processing of language, which requires more complex measures of presentation and assessment. For ease of use, all of the tasks are simply presented 'over the desk' and rely, for the most part, on assessment of accuracy rather than speed of responding. On the other hand, the materials we have provided allow for further investigation using techniques such as computer presentation and reaction-time measurement.

Construction of the tasks

Each test is designed so that the effect of one (or more) psycholinguistic variable is tested, while other variables that might also exert an effect are balanced across each of the experimental conditions. Thus, we might look at whether there is an effect of syllable length on word repetition by using one-, two- and three-syllable words, while matching them on factors such as word frequency, imageability and morphemic complexity. As far as possible, materials in each experimental condition are matched on a one-to-one basis, rather than matched as a group.

We used a number of resource books to construct the materials: word frequency ratings were taken from the Kucera and Francis (1982) word count; and imageability ratings were taken from the MRC Psycholinguistic database (Coltheart 1981). The 'Instructions for Use' that preface each task state the variables that are tested, and those on which the materials are matched though, for the sake of brevity, we have not included actual descriptive statistics.

In order to obtain reliable measures of a person's performance, and to allow accurate comparisons across conditions, a reasonable number of items have to be used. As far as possible one should aim to do the whole of a particular task; where applicable we have stated the conditions under which it may only be necessary to complete half of the task.

Data from non-brain-damaged subjects

The assessment tasks were given to 32 non-brain-damaged subjects. These subjects were generally the partners of the aphasics and were thereby loosely matched with them on age, educational and social variables. We have summarized their performance on each task in the 'Instructions for Use' section. In this publication we have provided only descriptive statistical analyses: means and
standard deviations (and ranges where relevant). We discuss what can be considered as 'abnormal performance' in the section on Interpretation of Test Scores, below. Note that some of the tasks were either newly compiled or changed substantially after we had gathered our data from non-brain-damaged subjects. In these cases we have been unable to provide normative data, and we recommend that you gather relevant control data for yourself before judging whether a particular pattern of performance can be considered to be impaired.

On the basis of the data from non-brain-damaged subjects we found that most of the tasks did not reveal differences between experimental conditions in terms of accuracy, and we had to make modifications to materials in just a few of the tasks. We cut down the Visual Lexical Decision task [25] to 15 items per condition because our subjects 'missed' some of the low-frequency–low-imageability words (because this did not happen when the same items were presented for Auditory Lexical Decision [5], it suggests that subjects only found written forms of these words unfamiliar). Subjects also had difficulty in writing low-frequency–low-imageability words to dictation [40], compared with those in other conditions—though, because of small numbers in this task to begin with, we have been unable to reduce the numbers further. We did, however, modify the task that examines oral reading or regular and exception words [35]. With our original set of 40 words per condition, subjects did show a significant regular word advantage in terms of accuracy. We therefore removed 10 exception words and their corresponding regular words, so that there are now 30 words per condition. Finally, our non-brain-damaged subjects found it hard to judge whether some of the mirror-reversed letters [18] were reversed or the right way round. We took out particularly difficult cases, together with the corresponding correct form, reducing the total number of items from 60 to 36. We have not retested our non-brain-damaged subjects with these revised materials.

Data from subjects with aphasia
The assessment tasks were also given to 25 subjects with aphasia. These subjects had all acquired aphasia after cerebrovascular disease. We placed no constraints on the time post-onset at which they were seen, or on the severity of language disturbance. We did, however, exclude subjects with severe perceptual disturbances and with severe dysarthria. We do not discuss patient data in this publication.

Using the battery
The battery is divided into four broad sections and each section is published as a separate manual: Auditory Processing; Reading and Spelling; Word and Picture Stimuli; and Sentence Processing. In each manual, all of the tasks requiring a particular response are grouped together (e.g. Minimal Pair Judgement, Oral Reading). This method or organization makes it easy to locate a particular task quickly. It also means that one can test for the effects of a number of linguistic variables (e.g. word frequency, imageability, syllable length) on a particular task in order to construct initial hypotheses about impaired and relatively spared language functions. One potential difficulty with this way of ordering is that tasks that are designed to work in tandem to test particular processing components (e.g. recognition and production of lexical morphology), or to assess the effects of particular linguistic variables (e.g. grammatical word class) are not necessarily close to one another in a manual, and may even be represented in different manuals. We have reduced this difficulty by pointing out relationships between tasks in the Instructions for Use that preface each task (we have more to say about this below).

Format of the tasks
Each task follows the same format. There is an Instructions for Use page that explains what the task is designed to test and how it was constructed. It also gives descriptive statistics from non-brain-damaged subjects, suggests where to go next in testing, and details any special points. Stimulus materials are presented next, followed by Presenter's Forms and Marking Forms. In tasks that are presented auditorily, stimulus materials are often given using the Presenter's Forms themselves. It is important to realize that each task can be marked using the Presenter's Form(s) alone. The Marking Form sets out the materials by condition, and shows how words are matched on a one-to-one basis across conditions. It requires the presenter to re-transcribe responses but, at the same time, it allows one to see at a glance particular response patterns.

A potential difficulty in having lots of different tasks and different forms for each task is that there are too many bits of paper to keep track of. We have tried to solve this problem in the following way. Each task has been uniquely identified by a number from [1] to [60]; for example, spelling-to-dictation of regular and exception words is task number [44], and spoken word–picture matching is task number [47]. Each form of each task is identified by that number, together with a separate page number and the total number of pages in the task. Thus, in the case of spoken word–picture matching, the Instruction for Use form is [47], page 1 (of 3).

Where to begin and where to go next
There is a temptation when assessing, say, particular skills involved in auditory processing, to start at task number [1] (Non-word Minimal Pair Judgements) and to work one's way through to number [17] (Phonological Segmentation of Final Sounds). Unless one has sound theoretical reasons to begin in this way, it is an inefficient strategy and one we would advise against. A good place to start an initial assessment of auditory processing is with spoken word–picture matching [47]. If a person performs poorly on this task, one can take account of what errors he or she makes and then compare ability on written word–picture matching [48], before going on to test auditory input processing skills. It is important to realize, however, that this is merely one suggestion for where to begin; it is one point of entry into the battery, rather than mandatory access. Above all, we want you to use the battery in a flexible way; flexible because it should be tailored to the requirements of an individual client. Similarly, although we have made suggestions about where to go next after each task, it must be emphasized that these are only suggestions.

Task administration
Most of the tasks are straightforward to administer, but most require explanation and practice to make sure that the person really knows what he or she is required to do. There are a number of tasks that require a 'same' or 'different' decision, but
conclusions about which processing components are operating at a sub-optimal level. However, it is not sufficient to examine the errors on one task alone, as different disorders can underlie the same symptom. Evidence must be accumulated on a number of tasks that are designed to tap different levels of processing.

Further assessments

The purpose of the PALPA battery is to allow one to derive hypotheses about the nature of the processing disorder in an individual with dysphasia. Its aim is to provide a firm grounding on which to base further assessments of a person’s difficulties and on which to plan directed treatment programmes. As such, PALPA tasks can be used in a flexible way—the materials provided in some of the tasks can be used with different procedures (computer presentation with timed responses, for example), or in different modalities, or to give baselines on which to assess the effects of particular treatment programmes—providing that the principles of experimental design and evaluation are respected.

Recommended further reading


References


Interpretation of test scores

In some cases, judging that performance is impaired is relatively straightforward. When the subject is required to make a choice, or some other form of categorial decision, estimates of what would be expected by chance are easy to make. For example, if a subject is given a binary choice of two responses (‘yes’ or ‘no’ in a same/different judgement task), then one would expect the subject to be correct, by chance alone, on 50% of occasions. Deciding whether performance differs from what would be expected by chance is then a simple statistical matter. However, in many cases a subject performs poorly on a particular task, but still manages to produce a substantial number of correct responses. The data that we have supplied from our non-brain-damaged subjects can help you to decide whether performance is impaired. Information about average (mean) number of correct responses and standard deviations allows one to calculate the number of standard deviations below the mean at which the patient scores. An arbitrary, but commonly used, criterion of two standard deviations (or further) below the mean can be judged as ‘impaired’.

One must be careful in assuming that a patient is performing ‘normally’ on a particular task if he or she falls within the range of our non-brain-damaged subjects. Remember that at least some of the tasks may be easy for a person without brain damage—their performance may be at ‘ceiling’ without taxing them in any way. Simply testing accuracy, rather than more rigorous investigation such as speed of responding, makes it difficult to conclude that an individual patient is performing a task without deficit, or to his or her pre-morbid level.

These methods of interpretation focus on number of errors. Another way to decide whether performance is impaired is to examine type of errors. One of the bases of the approach discussed here is to use the errors patients make to draw

with different criteria (e.g. whether two words are near-synonyms or whether they are unrelated; whether two words sound the same or slightly different), and, in these cases, it is best not to perform these tasks in close succession. When the explanation for a task may require extra elaboration, take account of the instructions in the ‘Special Points’ section of ‘Instructions for Use’.

In addition to the usual considerations when carrying out an assessment (such as avoiding eye-pointing to correct responses, and being careful to avoid intonational differences between items requiring ‘yes’ or ‘no’ responses), there are a number of others of particular importance when using PALPA. Many of the assessment materials involve repeated testing with the same materials, but in different modalities (albeit generally with different orders of presentation). To avoid practice effects or priming effects, care should therefore be taken not to use the same materials on more than one occasion on the same day. For similar reasons the person being assessed should not be told or made aware of which responses they made errors on, or which were correct, so that re-testing in a different modality or at a later date can be carried out (most people will accept this if the reasoning is explained to them).

The large number and range of tasks that make up the battery mean that one is generally able to select tasks that one is sure that the patient will be able to do with relative ease. Thus, it is useful during a test session (and perhaps to close it) to give a task that can bolster a patient’s confidence and to maintain morale.
The PALPA's proof is in the predicting

ROBERT T. WERTZ
VA Medical Center, Nashville, Tennessee, USA

Introduction
Another test for aphasia? Why? We have so many.

One reason for another test is that it does what existing tests do not do. Another reason is, of course, what it does do is important to do. The PALPA (Kay et al. 1992) may meet both criteria. The following is an impression, filtered through language barriers, bias, and experience, of whether the PALPA meets the criteria. Specifically, I ask what the PALPA appears to be, what it does not appear to be, and whether what it does justifies adding it to an aphasiologist's armamentarium. However, we might begin with what we expect an aphasia test to accomplish.

What aphasia tests do
We are no longer surprised that aphasiologists disagree. We do not agree on our definitions of aphasia (Darley 1982, Damasio 1981; our assessment of it (Byng et al. 1990, Goodglass 1990, Kertesz 1990, David 1990, Weniger 1990); or the way it should be managed (Chapay 1994). That is acceptable. Controversy does not deme us; it elevates us. And, as General George Patton observed, 'If everyone is thinking alike, no one is thinking'.

Specifically, what do we expect an aphasia test to do? Byng et al. (1990) say an aphasia test should 'first and foremost... elucidate the nature of the language impairment and indicate what aspects of language performance are most appropriate for treatment' (p. 67). They quarrel with existing aphasia tests for not doing either very well. Conversely, Goodglass (1990) says 'The objective is to provide the clinician with a guided comprehensive survey of language skills in a form that meets psychometric standards' (p. 94). David (1990) summarizes other possible purposes—screening; monitoring; predicting outcome (Lincoln 1988); consistent sampling of a range of patients and variety of disorders leading to a database of precise descriptions; gaining a meaningful understanding of the patient's problem from a theoretical approach to interpretation of test scores; comparison of affected versus unaffected skills with well known syndromes and over time; assistance of communication with other professionals (Crockett and Purves 1980) (p. 103). Rosenbek et al. (1989) add that an aphasia test should provide a measure of severity, differentiate aphasia from what it may masquerade as, provide a prognosis, assist in focusing treatment, and evaluate change over time. Thus, some want a test to accomplish some things, and others want a test to accomplish other things. Even those who have a similar list may arrange the list in a different hierarchy of preference.

Obviously, no single aphasia test will do what everyone wants, or even most of what some want. David (1990) observed that all aphasia tests are subject to some criticism. They should be, because all have weaknesses as well as strengths. Tests are tools designed to build some things and not others. Most of all, tests should be tests and not religious experiences (Wertz 1991). An ecumenical position is necessary, and doctrine depends on the specific purpose or purposes which appraisal is designed to achieve. Conversion and reconversion appear permissible depending on the patient's and clinician's needs at different points in time.

What the PALPA appears to be
PALPA is the test promised by Byng et al. (1990). It includes 60 assessments designed to identify language-processing problems in people who have acquired brain damage. Its purpose is to assist clinicians in formulating hypotheses about the nature of language-processing disorders in aphasic patients. In addition, it is designed to provide directions for additional appraisal and to plan appropriate treatment.

Model-based
A model of language processing is utilized to direct assessment of recognition, comprehension, and production of spoken and written words and sentences. The model consists of modules for processing, and it is assumed that brain damage can impair these modules selectively. Like most models of language processing, the PALPA's model consists of boxes and arrows. The boxes are assumed to store and process information. The arrows indicate communication among boxes. Three streams of processing are depicted—input of speech, pictures or seen objects, and print flow to output through speech or print. All streams may interact with a semantic system, and at some levels the streams interact with each other. Essentially, the PALPA's authors have asked how people communicate, and what is necessary to communicate the way people do. In addition, they have considered reports in the literature on what some brain-damaged patients can and cannot do, and they have adapted their model to accommodate these results. The product is a way of thinking about communication, the processing modules necessary, the pathways among modules, and which language-processing function depends upon each module or pathway.

The authors acknowledge their model is not universally accepted, and that use, and new information, may require modifications. Nevertheless, the model appears useful for identifying and interpreting different patterns of language impairment, guiding assessment, and assisting in focusing treatment. Certainly, those are some of the things which clinicians want aphasia tests to do; and, consistently, PALPA's 60 tasks are
designed to explore the integrity of the specific components and connections within the language-processing model.

The battery

Each test is designed as an exercise in controlling degrees of freedom. The influence of a psycholinguistic variable (for example, number of syllables in a word) is manipulated systematically; for example, in one- through three-syllable stimuli, while controlling for other influences such as word frequency, imageability, and morphemic complexity. This permits assessing the influence of a specific variable on patient performance and determining the influence of coexisting variables on performance. This degree of control is rare in existing tests for aphasia.

PALPA's 60 tests are grouped in four categories—17 assessments of auditory processing, 29 assessments of reading and spelling, eight assessments of word and picture semantics, and six assessments of sentence comprehension. Tests for each category are housed in separate manuals. Generally, similar tasks are grouped; for example, in auditory processing, four tasks utilizing minimal pair stimuli—non-word minimal pairs, word minimal pairs, word minimal pairs requiring written selection, and word minimal pairs requiring picture selection—appear in sequence. Some tests assess the integrity of a specific module in the modal; for example, Test 19, Upper-Case–Lower-Case Letter Matching, examines abstract letter identification. Other tests examine variables that may influence processing; for example, Test 24 looks at visual lexical decision with 'illegal' non-words, Test 25 looks at the influence of imageability and frequency on visual lexical decision, and Test 26 looks at the influence of morphology on visual lexical decision.

The materials

Each task, generally, follows the same format. An 'Instructions for Use' page precedes each task and includes a description of the task's purpose and what the patient is required to do; descriptive statistics for non-brain-damaged subjects' performance, when available; and suggestions for where to go next, including which additional assessments to administer if the patient does well and which to administer if the patient has difficulty. For some tasks, 'Special Points' are included to guide administration—how to orient the patient to the stimuli, how to present stimuli, and whether to use a repetition of stimuli. In addition, some 'Instructions for Use' pages for some tasks include literature references that elaborate the purpose of the task or support its use with aphasic people. Following the 'Instructions for Use' are the stimulus materials, a 'Presenter's Form', and a 'Marking Form'. Sometimes, depending on the task, the two forms are combined. Typically, the 'Presenter's Form' is used to record a patient's performance and, typically, each 'Presenter's Form' indicates the nature and/or description of each stimulus—for example, whether it is a regular or exception word and whether it is a word or non-word. The 'Marking Form' for this example groups stimuli into classes, e.g., regular words, exception words, and illegal non-words, for analysis. The 'Presenter's Form' permits calculation of the number of correct and incorrect responses, and the 'Marking Form' permits determination of the type of errors.

Normative data

Descriptive statistics are provided for most tasks. Typically, these are the mean correct and standard deviation for the task obtained by non-brain-damaged partners of aphasic subjects. Thus, there is loose comparability for age, education, and social status between the non-brain-damaged people providing the normative data and their aphasic counterparts. Normative data do not exist for all tasks, specifically, those that were constructed recently and those that have been modified. Tasks were administered to 25 aphasic people who were aphasic subsequent to cerebrovascular disease, but these data are not provided in the PALPA.

Administration

Considerable flexibility is permitted in where to begin testing and how to proceed with additional tests. The authors advise against starting with Test 1 and proceeding, in order, through Test 60. If one has an hypothesis about the nature of a patient's language disorder, I infer one can enter the battery and begin with a test designed to test the hypothesis. Utilizing the 'Suggestions for where to go next' for the test selected provides direction. If one does not have a specific hypothesis, the authors suggest beginning with assessment of auditory processing by administering Test 47, Spoken Word–Picture Matching. If performance is poor, that performance is analysed, and Test 48, Written Word–Picture Matching, is administered before proceeding to assess auditory processing skills. Users are cautioned not to administer some tests in the same session, because similarity in stimuli may influence performance.

Patient performance

What does a patient's performance on the PALPA mean? What constitutes impairment? The authors suggest utilizing data from the non-brain-damaged subjects. A suggested arbitrary decision to indicate impairment is performance two standard deviations below the non-brain-damaged subjects' mean. However, the user is cautioned to consider 'chance' performance on binary—'yes' or 'no'—decision tasks. Moreover, the authors suggest going beyond accuracy and considering type of errors, speed of responding, and assumed pre-morbid ability. Consistent with the PALPA's rationale, the user is also urged to compare performance on different tasks designed to assess different levels of processing.

The PALPA is

What does the PALPA appear to be? Generally, it appears to be exactly what it reports to be. A variety of tasks are provided to permit inferences about the nature of an aphasic patient's processing. This is done by comparing PALPA performance with the processing model on which the PALPA is based. This comparison permits constructing an hypothesis about treatment, developing treatment tasks consistent with the hypothesis, administering them, and evaluating the results to accept or reject the hypothesis. If all of this works, it brings order and consistency to a practice that, unfortunately, is often chaotic.
What the PALPA does not appear to be

As indicated above, an aphasia test will not satisfy everyone. This results, in part, from different purposes for testing and, in part, from different criteria for tests. I have discussed what the PALPA seems to me to be and do. Now let us consider what the PALPA seems not to be or do.

Psychometrics

The PALPA is not standardized or, at least, fully standardized. The authors acknowledge this. However, lack of standardization presents difficulties. It permits interpretive errors—attributing inappropriate meaning to performance, variable errors—differences between testing at one time and testing at another and differences among those who are tested, personal errors—differences within and among observers, and constant errors—assuming the test is measuring something that it does not.

Standardization assists in reducing interpretive errors. Reliability reduces variable errors. Objectivity reduces personal errors; and validity indicates that the measure does what we think it does and, therefore, is not influenced by constant error. Of course, a test may be highly standardized and highly unreliable, lacking in objectivity, and completely invalid. Conversely, a highly reliable test may not have been well standardized and may be completely useless for its stated purpose, because it is not valid. Objectivity, however, will always be reflected in reliability. Ideally, a test could be judged solely in terms of its validity, but if it is going to be used for a variety of purposes—to indicate severity, differentiate among disorders, focus treatment, measure change, etc.—it must be objective and reliable. Standardization assists in developing both. What is the PALPA?

The authors say the data we have collected so far are not intended as a full standardization of the battery. Thus, we have not carried out psychometrically satisfactory measures of validity or reliability. We have published the tasks because of the great demand for well-controlled psycholinguistic materials. So, the PALPA is not fully standardized and does not have demonstrated validity and reliability. Let us consider each, beginning with standardization.

To demonstrate validity and reliability, one typically begins with standardization. If the instructions, stimuli, and responses are specified precisely, then the test can be administered and scored in exactly the same way to the appropriate sample or samples to test its validity and reliability. Variance in administration and scoring may or may not influence validity and reliability, but we do not know, so we begin with invariant administration and scoring, essentially, standardization. Thus, standardization is separate from validity and reliability even though it may influence both.

Given the authors' rationale and instructions for administration, I wonder if the PALPA can be standardized without a change in both. For example, permitting the tester to enter the battery by administering any subtest, not administering all 60 subtests in a specified order, and not administering the similar subtests during the same session all argue against standardization. Moreover, in some tasks the authors indicate it may be necessary to present only half of the stimuli. Thus, when a specific standard is not prescribed, it is difficult to achieve standardization. Is this a problem for the PALPA? It may be if lack of standardization obviates demonstrating validity and reliability.

Similar problems exist in demonstrating that the PALPA is valid. We assume a test is valid if it measures what we say it measures. This can be done in a variety of ways, depending on the type of validity—predictive, concurrent, content, construct—one is interested in demonstrating. The PALPA's purposes are to derive hypotheses about the nature of the processing disorder in an aphasic patient, direct further assessment, and plan appropriate treatment. Thus, it seems necessary to demonstrate each type of validity: predictive to predict future performance, concurrent to estimate present performance, content to determine the test's estimate of performance on an important set of tasks, and construct to explain performance psychologically.

Two difficulties seem apparent. First, the PALPA's purposes utilize constructs—'nature' and 'processing'—that are difficult to define, quantify, and demonstrate empirically. Second, the PALPA was developed because a similar measure does not exist. Thus, the typical approach—validating a test for aphasia by comparing performance on it with performance on an existing, apparently valid test for aphasia—is not a straightforward option. Certainly, it will not be straightforward or easy to demonstrate the PALPA's validity.

Demonstrating reliability is equally problematic. Usually, reliability is assisted by standardization. Without standardization—a prescribed method and order for presenting and scoring—it will be necessary to demonstrate reliability for the infinite variety of presentation orders and the variety of methods permitted. For example, test–retest reliability can be established for a specific PALPA subtest, but will that reliability differ if the specific subtest is preceded by different subtests? The authors fear it may, because they caution the user not to administer some subtests in the same session.

The kinds of reliability one might desire are test–retest reliability, intra-examiner reliability, and inter-examiner reliability. The first, test–retest, is necessary to demonstrate stability in a patient's performance. In the PALPA, we do not want to infer a patient has processing difficulty in one module, or performance is affected by one psycholinguistic variable today, and discover neither appears impaired tomorrow. Moreover, if we elect to treat a patient and use the PALPA as a change measure, we want it to reflect change resulting from the treatment, or time, or both, and not test-to-test fluctuation in performance. The second type of reliability, intra-examiner, is necessary to demonstrate stability in how behaviour is observed by the same clinician. If patient performance is stable, we want a clinician's observations to indicate that it is. For the PALPA, this is important for stating severity, describing the nature of the processing deficits, selecting additional assessments, creating a hypothesis about treatment, and measuring change in patient performance as a result of treatment, time, or both. Finally, inter-examiner reliability is demonstrated when there is agreement between or among two or more examiners. Essentially, when we administer the PALPA, you should see what I see, and both of us should see what everyone else sees. If we cannot demonstrate inter-examiner reliability, patient performance varies as a function of the examiner and not as a function of the patient. Thus, different clinicians observe different severity and different processing deficits, select different additional assessments, and create different hypotheses for treatment for the same patient.

Of course, if normative data are reported on non-brain-damaged people's performance, we will want the same kinds of reliability for their performance. Currently, none of these types of reliability has been reported for normal or aphasic performance on the PALPA.
Normative data

Generally, we like an aphasia test to indicate the presence of aphasia. One way of doing this is to administer the test to ‘normals’—people who are not aphasic—and compare performance with people who are aphasic. Not all authors of aphasia tests have done this, and those who have realized the problems with the procedure. For example, the 93-8 cut-off score to indicate ‘normal’ performance on the Western Aphasia Battery (WAB) (Kertesz 1982) is based on Kertesz’s normative sample. All subjects obtained aphasia quotients (AQ) of 93-8 or above. However, some aphasic people also obtain an AQ of 93-8 or above. Similarly, assumed ‘normal’ performance on the Porch Index of Communicative Ability (PICA) (Porch 1967) is 15.0. However, Duffy and Keith (1980) have demonstrated that some ‘normals’ score below 15.0 on the PIAC, and ‘normal’ performance is related to education and age.

The PALPA’s authors have provided some ‘normative’ data for 44 of the 60 subtests. These are presented as means and standard deviations. Sample size varies among subtests. Certainly, this is a beginning, but more data are necessary. The way the data are presented also requires attention. Given the range and variety of the PALPA’s subtests it will be necessary to be very precise in establishing ‘normal’—non-aphasic—performance for each subtest. Without this information, we will not be able to achieve the PALPA’s purposes—to describe the nature of the language processing deficit, select additional appraisal measures, and focus treatment. Essentially, we need sensitivity and specificity data. Few tests for aphasia provide this. All could, including the PALPA.

Diagnosis

Should an aphasia test differentiate aphasic people from those who are not aphasic? I have implied above that an aphasia test should separate aphasic performance from ‘normal’ performance. The PALPA’s authors agree, as indicated by their presentation of data on non-brain-damaged individuals. However, some (Darley 1982, Holland 1982, Wertz 1982) suggest all brain damage does not result in aphasia. The problem, of course, is defining what we call aphasia. Au et al. (1988) and Appell et al. (1982), for example, suggest communication impairment in dementia is aphasic. Thus, for them, it is not important to separate aphasic performance from demented performance, because the two are the same. However, for purposes of diagnosis, prognosis, and treatment planning, it seems important to determine who is aphasic and who is not.

Does the PALPA differentiate? We do not know. Some of the sequelae subsequent to dementia, traumatic brain injury, right hemisphere brain damage, apraxia of speech, and dysarthria may imply language-processing deficits when filtered through the PALPA and compared with the model on which it is based. Will the pattern of performance indicate who’s who and who’s not? It would be very useful if it did, because different disorders, assuming that different disorders exist, may demand different explanations about the nature of the language-processing deficits; require different additional assessments; and different hypotheses about treatment if, in fact, treatment is an option for some disorders. Asking the PALPA to differentiate may be unfair. No existing test does this very well. The PIAC attempts to, but its efforts are limited in application. Nevertheless, it is important to try, and the PALPA’s rationale, purposes, and organization may succeed where other assessments have failed.

Comprehensiveness

The PALPA does not assess every bit and piece of language that may be impaired subsequent to brain damage. No test for aphasia does; but how many bits and pieces are necessary to achieve the PALPA’s purposes? The authors indicate they have not considered everything. The PALPA ‘provides materials for only a circumscribed part of what needs to be assessed when one investigates the language abilities of a person with dysphasia’ (Kay et al. 1992: p. 1). Some areas—sentence production, inference-making, ‘on-line’ processing of language—are not assessed.

Certainly, the PALPA has depth, but it may lack breadth. No existing measure probes auditory processing, reading and spelling, picture and word semantics, and sentence comprehension as thoroughly or systematically. Similarly, none exerts similar control for psycholinguistic variables or permits similar inferences about their effects on performance. Auditory comprehension, reading, naming, repetition, and writing tasks are included. However, the PALPA does not include measures of conversation to assess information content, fluency, and grammatical complexity. Similarly, assessment of writing is limited, and measures of reading are limited to words and sentences. Thus, the PALPA does not do what some tests do. Even 60 tasks will not do everything. More important, however, is whether the PALPA is sufficiently comprehensive to achieve its purposes—to derive hypotheses about the nature of the processing disorder in an aphasic patient, suggest the need for additional assessments, and focus treatment. Its depth and breadth appear appropriate to do these for some patients, but it may be too shallow and too narrow to accomplish its purposes for others.

User-friendliness

Cronbach (1960) observed: ‘preparing a good manual is difficult’ (p. 101). He suggests a test manual must be sufficiently clear for the qualified to comprehend and, conversely, for the unqualified to realize he or she is, indeed, unqualified. The danger is not in not knowing, but in not knowing one does not know. Does the PALPA meet these criteria?

Patient performance on PALPA subtests should permit the tester to achieve the test’s purpose—to infer the nature of the patient’s processing disorder, decide what to assess next, and formulate an hypothesis about treatment. The latter depends on the two former. Additional assessments are, typically, suggested in the ‘Suggestions for where to go next’ that accompanies each subtest. Making an inference about the nature of the processing disorder, based on a patient’s performance on a subtest or subtests, is, for me, not always straightforward. Some subtests are specifically related to modules in the model; for example, subtest 1 assesses auditory phonological analysis. Others are not specifically related to the model; for example, subtest 13 assesses auditory digit span, and the authors indicate performance is related to ‘some of the components of the “phonological short-term memory system”. I do not find “phonological short-term memory system” in the model. Is this in the “phonological input buffer”? Probably, but ease of use would be promoted if the authors, in their ‘Instructions for Use’, systematically and consistently indicated how each subtest related to a module in the model or the interaction among modules. Even more user-friendly would be a section in the manual that related each of the 60 subtests to specific modules or interactions among modules in the model. It is difficult to make things ‘foolproof’, because fools are so ingenious, but one should try.
Similarly, one of the PALPA’s strengths is its systematic manipulation and control of psycholinguistic variables. Typically, these are indicated by bold-faced abbreviations in each subtest, and usually they are defined in each subtest’s ‘Presenter’s Form’ and ‘Marking Form’. Usually, but not always. Sometimes one can infer what an abbreviation indicates, for example, ‘A’ = adjective, and sometimes one cannot. Moreover, sometimes the abbreviations change; for example, ‘E’ = exception but, later, ‘e’ = exception. Ease of use might be improved if abbreviations were consistent, and all were contained in a glossary for reference.

No test is error-free, especially in its first edition. The PALPA is no exception. For example, subtest 59 tests ‘Written Comprehension of Locative Relations’. It is similar to subtest 58, ‘Auditory Comprehension of Locative Relations’. In fact, it may be too similar. The task in subtest 59 is to assess comprehension of locative prepositions in written phrases, but there are no written phrase stimuli. The ‘Presenter’s and Marking Form’ instructions are ‘I’m going to say a phrase. I want you to point to the picture that fits.’ These instructions are identical to those in subtest 58. Either something is missing, or I am missing something. Additionally, in Nashville, Tennessee, ‘onion’ and ‘union’ are two-syllable words. In subtest 7 they are listed as three-syllable words.

The PALPA was developed in the United Kingdom. Can it be used elsewhere? It can with some modifications. For example, in subtest 4, ‘tack’ may be perceived as a ‘check’ in the USA, and ‘tip’ in the USA is a ‘dump’. Similarly, in subtest 47, ‘hosepipe’ is a ‘nozzle’ or ‘hose’ in US yards. In subtest 40, only Dan Quayle might spell ‘valour’ as ‘valour’ in the USA. Incidentally, there are no instructions on the subtest 40 ‘Presenter’s Form’. I suspect they are the same as those on subtest 39, ‘I’m going to say a word to you. Then I want you to write it down’. A few of the phonetic symbols used to indicate ‘non-word minimal pairs’ in subtest 1 are not common on this side of the Atlantic. These differences are fun and educational. However, the PALPA may need modification when used in the colonies.

Proof is in the predicting

It is difficult to decide when to release one’s work for peer perusal and use. Schuell (1965) spent 18 years constructing the Minnesota Test for Differential Diagnosis of Aphasia before putting it in print for the public. The PALPA’s authors have elected to release it now. There is merit in both approaches. Early release, however, usually produces more promise than proof. That is acceptable. Use can clarify and question; it can also provide direction. The work necessary on the PALPA, I believe, is three-fold: correct the obvious errors, provide the psychometric base, and test the ability to predict what is promised. The first is easy; the second and third take time to collect the data.

We want our measures to be psychometrically sound. If they are not, we are never certain why they do or do not work. Standardization provides consistency, validity provides comfort, and reliability provides confidence. The PALPA lacks all three. Without standardization we do not know why our results differ from yours, or even if we agree, why we agree. Without validity we do not know whether what we think we are doing is really what we are doing. Without reliability we are limited in application and interpretation. The PALPA is ‘hypothesis-driven’. That is good, because an hypothesis can be tested. The test, however, requires rigour, and it requires precision. The PALPA’s lack of standardization, validity, and reliability are not fatal flaws. They are the work to be done. Psychometrics may be rare in psycholinguistics, but they are essential.

Will the PALPA do what its author’s predict it will do—permit hypotheses about the nature of processing disorders in aphasia, direct selection of additional assessment, and permit selection of appropriate treatment? We do not know. These are the PALPA’s promises, and we seek proof. Words such as ‘nature’ and ‘processing’ create angst. They are employed when we cannot be more precise. The PALPA’s rationale, organization, and content may permit precision. If it does, a considerable contribution will result. If the PALPA is able to predict the ‘nature’ of an aphasic patient’s ‘processing’ disorder, we will be able to answer specific questions—What do we mean? How do we know? And what difference does it make?—when we talk about the ‘nature’ of an aphasic patient’s ‘processing’ disorder. Similarly, if the PALPA can predict additional assessments necessary to explain a patient’s ‘processing’ problems, it will assist in reducing our active uncertainty about the ‘nature’ of these ‘processing’ problems. That too will be a considerable contribution. Not knowing is a problem; not knowing how to find out is a real problem. Finally, if the PALPA can predict a plan for efficacious management, clinicians benefit and patients prosper. We find it difficult to explain and justify much of what we do when we say we do therapy (Holland 1992). The PALPA may assist us in doing what, currently, we cannot. These are the PALPA’s promises, and it seems to me that the promises need to be proven by empirically based, predictive precision.

Is the PALPA a revolution in aphasia testing? I think not. Revolutions typically demolish what is, disrupt the normal flow of progress, and destroy the best elements of the past. Perhaps the PALPA is an evolution and a liberation in aphasia testing. It does not undermine traditional strength as revolutions tend to. Instead, it appears to have evolved from the author’s experience and from clinical evidence. It may liberate thought and, consequently, permit evolutionary development in aphasia testing. Moreover, the PALPA may transcend the merely academic and permit academic practice. If one is dealing only with theory, things may seem much more complicated than they are. If one has spent a few hundred hours sitting knee-to-knee with aphasic people, one knows the role of theory in academic practice (J. C. Rosenbek, personal communication, 1994), and things seem only as complicated as they are. The PALPA appears to have resulted from considerable time in the knee-to-knee position.

However, I do worry. My raccoon approach to an aphasia test is to take it and to give it. One of my colleagues administered the PALPA to me, and I administered it to one of our patients. I did better than the patient. In fact, where norms are available, I performed within the normal range. This is troublesome, because many suggest my hold on normality is, at best, questionable.

References


For many years evaluation of aphasic patients was based on observation of what tasks the patient can perform, whether he or she can read, write, repeat, name, etc. Comparison of performance on such tasks classified the patients in one of the main aphasic clinical syndromes. In the context of fluent speech, for instance, preserved repetition and severely impaired auditory comprehension led to the diagnosis of transcortical sensory aphasia, and the reverse pattern (relatively spared auditory comprehension and severely impaired repetition) led to the diagnosis of conduction aphasia. However, it is not the task (repetition, comprehension, etc.) that is represented in the brain, but the cognitive processes necessary to perform it; inability or difficulty in the execution of the task are the observable symptoms.

These recent decades have witnessed the rapid expansion of cognitive neuropsychology based on the assumption that language (as well as all other cognitive systems) has a complex architecture consisting of many functionally independent components or modules. The application to the study of aphasia of the approach proposed by cognitive neuropsychology has led to substantial changes. The focus of research has shifted from observation of symptom complexes to the in-depth study of the organization of pathological language. Since pathology is a deviation from the normal, normal processing must be known and the pathological behaviour is compared to normal processing in order to identify the functional damage.

The study of brain-damaged patients within the cognitive neuropsychological approach has been highly stimulating and successful. Caramazza and Badecker (1991: p. 211) have argued that 'one is compelled to use single-patient methodology whenever one wishes to employ data from acquired cognitive impairments to motivate specific claims about normal processing mechanisms'. The case study began to be considered the most powerful empirical procedure for drawing inferences about normal function, and results from single case studies are considered to produce strong evidence for discriminating between theories of normal function. The cognitive neuropsychological literature is rich in in-depth studies of single cases (see, for instance, Howard and Franklin 1987). Generally, tasks have been devised individually and compiled according to what aspects of language had to be studied in that patient. This obviously necessitates much time and ingenuity on the part of the examiner, and is not possible in a clinical setting where patients are referred for rehabilitation. Speech therapists can rely on such tests as the Boston Diagnostic Aphasia Examination (BDAE), the Porch Index of Communicative Ability (PICA), or the Aachener Aphasia Test (AAT) that have undeniable merits and allow clinical diagnosis, but not identification of the functional locus of damage (such as the phonological output lexicon or the sublexical reading routine, for instance) because they are not based on detailed models of normal language processing.

The PALPA responds to a great demand for well controlled psycholinguistic material, and it beautifully fills this gap. The stimuli are carefully chosen according to linguistic parameters, such as frequency of use, length, regularity, etc. The test is easy to understand and instructions are clear. The authors regret that there is still no full standardization of the battery. For my part I do not think that a standardization would be of much help, although it would be interesting to know normal performance on these tasks. The test is not devised to discriminate pathological from normal performance, as is the Token Test (De Renzi and Faglioni 1978), for instance. PALPA has been devised to give a qualitative analysis of the patient's behaviour necessary to identify the functional damage.

Notwithstanding its obvious and unquestionable merits, PALPA is, however, not

PALPA: an appreciation and a few criticisms

ANNA BASSO
Università di Milano, Milan, Italy

Evaluation of such a complex phenomenon as aphasia poses many theoretical problems, the most important of which refers to definition of the object of study, in this case aphasia. The test used will differ according to the definition of aphasia used, since the pattern of included and excluded tasks depends on what is considered important. Results will only confirm or fail to confirm expectations, and only exceptionally will something totally new and unexpected be found with reference to the definition given. To put it another way, results depend on the methodology adopted.
perfect. One obvious limitation is the imbalance between assessment of single-word processing and assessment of sentence processing. In the Introduction the authors note that their initial aim was to construct a battery for evaluation of acquired reading and spelling disorders (Kay et al. 1992: p. 2), which probably explains why reading and spelling assessment are so detailed and thorough. The imbalance between the attention given to single-word and sentence processing can partly be explained by the present state of knowledge. Models of the lexical–semantic system are highly elaborated and in general agreement, though not entirely identical. The same cannot be said for models of sentence processing, which are far less elaborated and much more heterogeneous.

While this can partly account for the aforementioned imbalance, the capacity for producing and comprehending sentences is the most exquisite human characteristic of language, and distinguishes it from other systems of communication. As such it deserves thorough examination. In default of a sufficiently detailed and agreed-on theory of mechanisms involved in production and comprehension of grammatical structures, the observed behaviour in sentence production and comprehension of aphasic patients can be used as a guideline.

PALPA evaluates auditory (and written) sentence comprehension and care is taken to test different types of sentences: reversible, passive, etc. However, the ability to judge the syntactic well formedness of a sentence is independent of the ability to understand the sentence, as shown by patients who cannot point to the correct picture but perform much better in a grammaticality judgement task (Linebarger et al. 1983, Lukatea et al. 1988, Shankweiler et al. 1989). For remediation planning it is important to know whether a patient with sentence comprehension deficit is or is not sensitive to syntactic structure well formedness. A grammaticality judgement task would have answered the question.

Moreover, although damage to the syntactic and grammatical mechanisms of language may be limited to language comprehension, in the majority of cases both sentence comprehension and production are involved, and in some rare cases only production. PALPA, however, does not provide tasks for the evaluation of sentence production, although probably more is known about sentence production deficits than sentence comprehension deficits. Controlled picture description tasks can be used to evaluate various grammatical constructions, such as active and passive sentences, and could fill the gap.

Another important deficiency is in verb retrieval. Tasks 53 and 54 investigate the patient's capacity to retrieve nouns, but no task requiring the patient to retrieve action names is included. Many agrammatic patients show structural simplification in sentence production, and Berndt (1991) suggests that a problem in realizing verbs for production can be an important component of the patient's failure to encode more complex sentences because the verb encodes much information about the form a sentence will eventually take. Even if this were not the case, an action-naming task would be important, since some (agrammatic) patients have a specific deficit of verb retrieval, whereas other types of patient (anomic aphasic) are better able to provide verbs than nouns. To know whether retrieval of nouns and verbs is equally damaged or not would be helpful for better characterization of the functional damage and for rehabilitation planning.

In conclusion, PALPA represents a step forward in aphasia assessment, allowing as it does for in-depth psycholinguistic analysis of the language disorder. However, as it stands, and even though it is stated that the battery is not to be used in its entirety.

References


Address correspondence to: Anna Basso, Istituto di Clinica Neurologica, Università di Milano, Via F. Sforza 35, 20122 Milan, Italy.

The PALPA: a valid investigation of language?

ALISON FERGUSON and ELIZABETH ARMSTRONG†

Lidcombe Hospital, Sydney, Australia
†Macquarie University, Sydney, Australia

Introduction

The long-awaited PALPA by Kay et al. (1992) is all that it claims to be, but no more than this. As a set of resources for clinicians already familiar with its cognitive neuropsychological theoretical background it provides controlled stimuli and tasks which allow for systematic replication of case studies, baselining for individual cases, and for making comparisons with appropriate control subjects (as recommended by Shallice in 1979). As the first major contribution to aphasia assessment for a number
of years it represents a major landmark in the clinical field. However, both the excitement we have felt in awaiting its arrival, and its relative novelty, should not, we feel, obscure the observation that there are important restrictions on the validity and usefulness of the PALPA. In this paper we have focused on these restrictions, firstly considering those which relate to information processing, and secondly, considering those which relate to our fundamental understanding of language.

Information processing

The strengths of using the information processing approach to assessment proposed in the PALPA are often held to be:

1. its theoretical basis,
2. its systematic, hypothesis-forming approach,
3. its ability to generate therapy.

Clinicians need to be aware that the PALPA strives for a broad understanding of what is described as the 'functional architecture' of the language-processing system, and that this approach has been criticized for failing to increase our knowledge of how language is stored, accessed, and processed (Seidenberg 1988). Clinically, this theoretical restriction to the PALPA results in the testing procedure ignoring the significance of observable behaviours such as delay, rehearsal and repair, scoring for accuracy alone. In addition, the processing model proposed by the authors very much reflects a psychological view of language as a mental process like other mental processes, rather than reflecting a uniquely linguistic viewpoint. Like previously developed aphasia test batteries (Boston Diagnostic Aphasia Examination—Goodglass and Kaplan 1972, Western Aphasia Battery—Kertesz 1982, Neurosensory Center Comprehensive Examination for Aphasia—Spreeen and Benton 1977) it focuses on specific metalinguistic behaviours such as repetition, naming and lexical decision rather than on the linguistic knowledge available to speakers of the language which is dependent on a grammar incorporating detailed notions of semantics and syntax. As a result, from a linguistic perspective, the view of language presented is sufficiently restricted as to jeopardize its validity.

The hypothesis-forming approach made possible through the schematic display of a componential view of language has been highly attractive to clinicians seeking a systematic approach to moving from assessment to therapy. At first it was with some disappointment then that the PALPA was found to give too little direction for such a hypothesis-forming approach, and in particular that such scant attention was paid to the potential 'entry points' to selecting tests from among this large battery. Clinicians are advised against starting at the beginning on the Auditory Processing section, with the authors suggesting that 'spoken word—picture matching is a good place to start'. One is then cautioned that 'we want you to use the battery in a flexible way, flexible because it should be tailored to the requirements of the individual client' (p. 20). However, one is given no theoretical reasons to begin anywhere in the battery. Presumably the clinician is supposed to have an intuitive idea as to the problems which the client has—either from assessing performance in functional situations or performing a standardized aphasia battery which then functions as a screening test. The lack of direction in this area presupposes that the clinician has a better way than the PALPA of detecting impairments in the first place. If one is supposed to assess from functional status, then some connection between the two should be incorporated into

theoretical notions of the PALPA, and certainly incorporated into the model at some point. If, on the other hand, another aphasia test battery is sufficiently sensitive to pick up hints of impairments, then the clinician should be advised which ones the authors think the most suitable in relation to the PALPA.

In terms of direction for treatment, while therapy content can be derived from the results of testing, the theoretical model proposed provides no guidance for whether the therapy is to involve stimulation, re-learning or compensation, for example. While the test developers see such treatment decisions as being the province of the clinicians rather than the test, the authors' inability to incorporate such decisions into the theoretical base to deal with the theoretical base to deal with such fundamental issues of information processing as learning, illustrates, in our opinion, a substantial restriction in the usefulness of this theoretical base to clinical work.

View of language

The information-processing model on which the PALPA is based is clearly presented in the Introduction. However, it unfortunately fails to present a sufficiently detailed view of the linguistic information being processed to be compatible with current theoretical notions about language.

The PALPA's focus on the single word as the main item of analysis perpetuates in aphasiology the rather outdated view (as far as linguistics is concerned) that investigation of the word as an isolated entity will reveal significant insights into the aphasic speaker's overall language competence or functional performance. While sentence-processing is included (to be commented upon below), it is disappointing that no aspects of discourse are even mentioned in the Introduction. The authors largely dismiss sociolinguistic views of language, with conversational analysis described as being at 'an early stage of development' (Introduction, p. 1), which is surprising since this field of endeavour has been in existence for some 20 years now. It is even more surprising that discourse analysis and discourse processing do not rate a mention, given the extensive research which has been done in the area, related to aphasia in particular. (For overviews of these areas see Davis and Wilcox 1985, Lesser and Milroy 1995 and Joanne and Brownell 1990.)

Yet, even taking the single-word focus as a reflection of the arrangement of the mental lexicon as debated in current literature, the model proposed is far from adequate. Grodzinsky (1990) has commented on the 'pre-theoretical' nature of information-processing models such as the one under discussion, making the point that, even if focusing on lexical aspects alone, the structure of the lexicon is not sufficiently detailed in these models to make claims regarding linguistic processing which would relate to any normal theory. For example, he quotes the argument in linguistics in which the extent to which the lexicon participates in determining sentence structure is debated, i.e. some linguists see the lexicon as central while others see its role as less important. The model proposed by Kay et al. leaves their perspective on the lexicon in such terms far from clear, and hence leaves interpretation of findings from such investigations as provided in, for example, the Picture and Word Semantics section of the PALPA, limited at best in terms of linguistic theory.

The authors note that they have 'not chosen to assess many important aspects of language use, such as inference-making' (p. 18), saying that other aspects can be 'added on' in the future as seen fit. However, this again reinforces the simplistic view of language presented in such a model. Even in a constituency model of language—
which one accepts that morphemes make up words, words make up sentences, etc.—it is not really acceptable to discuss only one aspect of the language system in isolation if one is attempting to characterize the language abilities of a speaker. While the PALPA does include a sentence-processing section, the theoretical principles outlined in this section appear to have little connection with the single-word models outlined in other sections, and in the summary of the model given in the Introduction. It is unclear where sentence processing fits into the model described at all. This is a major problem with the whole framework, as the authors appear to be saying that modules incorporating different levels of language functioning can work on differing linguistic principles. No doubt if discourse processing was added at some time in the future, further different principles would also be incorporated into the model to accommodate this ‘higher’ level. There appears to be no underlying model of language being utilized which can be applied to all levels of linguistic functioning. Hence, without an exhaustive linguistic theory as its basis, the model allows very few conclusions to be reached about what is actually being processed, and serious discussion about the individual’s language system is therefore limited.

A modular view of language views language as a primarily psychological entity existing as a system within each individual. In this way language is seen as fairly much a static set of rules and regulations which an individual uses for a variety of purposes. A sociolinguistic perspective, on the other hand, views language as a set of resources through which meanings can be exchanged in a dynamic way. While an individual obviously has to have a certain capacity within his/her mental apparatus to deal with language, the way the language is organized within that mental faculty has a great deal to do with the way language is used by the individual. Hence, factors such as context are relevant in the language-processing system itself—they are not merely extraneous conditions which exist outside the language system. They are driving forces within the organization of that system, and to ignore them means ignoring integral parts of the system itself. The notions of form and function in language have long been artificially separated, as Kay et al. acknowledge in the Introduction. This has largely been a reflection of the adoption of Chomsky’s dichotomy of competence/performance, but future tests may do well to take other theoretical approaches into account which recognize the fundamental inseparability of these two important aspects of language.

To conclude, we welcome the arrival of the PALPA, recognizing its importance to aphasiologists interested in the cognitive neuropsychological approach to assessment. However, we would urge clinicians to recognize the inherent limitations of this approach to both assessment and treatment.

**References**


---

**The PALPA: a commentary and consideration of the clinical implications**

JANE MARSHALL

City University, London, UK

Many commentators have become dissatisfied with the standard aphasia test batteries. Some have called for more functionally relevant assessments (Prattall 1992), while others have demanded alternative psycholinguistic measures, which can identify the locus of impairment on a model of language processing (Byng et al. 1990). The much-longed-for PALPA responds to the second demand.

This mammoth and invaluable resource provides clinicians with 60 language assessments, encompassing both single-word and sentence-level skills. Each test carefully controls the relevant psycholinguistic parameters, so that conclusions can be drawn about the functioning of the patient’s language system. Clinicians are assisted in this process by well-designed score sheets, which typically isolate the relevant factors for comparison. The tests are not simply diagnostic. In most cases the number of items is also generous enough to provide pre- and post-therapy evaluation measures, both of the treated functions and, if desired, untreated control tasks.

Criticisms of new tests are always possible. I found the organization of the collection a little eccentric. The authors seem unable to decide whether to group tests under modality headings, such as ‘Reading and Spelling’, or the relevant area of the processing system, such as ‘Picture and Word Semantics’. The location of some individual tests was also surprising; for example it is odd to find tests of repetition, which involve output, as much as input, in the ‘Auditory Processing’ section. Organizing a collection of this kind is obviously very difficult, given that the relationship between each test and the different components of the language system is by no means one-to-one. The authors have tried to overcome some of the difficulties by outlining which language skills are tested by each measure, and highlighting the connections between the different tests. Some form of cross-referenced index might also have been helpful.

More detailed gripes are possible. Since when was ‘Chad’ a non-word? (Homophone Decision Test 28) and I defy anyone to recognize the picture either of the ‘bead’ or ‘rice’ in the Minimal Pair Test 4. (The picture of ‘gut’ in the same test is also downright revolting.) The Rhyme Judgement Test (14) depends on very restricted mappings.
between words and pictures. A patient could fail simply because he or she has an
alternative name in mind, e.g. ‘book’ instead of ‘page’ (item 4), ‘deer’ instead of ‘fawn’
(item 16), etc. Yet these are quibbles. Therapists know that all tests have wayward items
and will interpret the scores accordingly.

I had some more profound difficulties with aspects of the Sentence Comprehension
section. Unlike the lexical assessments these tests are not explicitly related to any
model of sentence processing, which of course partly reflects our stage of thinking in
this area. However, clinicians are expected to apply quite sophisticated linguistic
concepts, such as ‘trace theory’. The demands imposed on patients are equally severe.
For example the Sentence–Picture Matching tests (55 and 56) require subtle picture
interpretation skills, as well as high-level semantic/syntactic abilities (e.g. ‘the horse’s
likely to kick’ test 55 item 13). For many patients these demands may be exacerbated
by the sheer peculiarity of some of the depicted events (e.g. ‘The horse is carrying
the cat’). It also seems to me that the test of verb and adjective comprehension
(no. 57), which aims to be a lexical assessment, in fact requires sentence-level skills
in order to interpret the definitions.

A more general charge would be to suggest that the PALPA expects too much of the
practising clinician. It has been said that language assessments are only as good as
the person using them, and this has never been truer than in the case of the PALPA.
As stated in the Introduction, one minimal requirement for using these tests is an
understanding of the language model on which they are based. But there are more
extensive demands. The PALPA asks the clinician to adopt an experimental approach
to the assessment of patients. Some reasonably specific hypotheses must be formed
about the nature of the patient’s language impairments before the PALPA is taken out
of the cupboard. The clinician must decide which assessments to choose in order to
test these hypotheses, and understand how to interpret the data which they supply.
Above all the PALPA demands an ability to conflate results from different assessments
in order to deduce the underlying functioning of the language system. This collection
is, therefore, more than just a few new tests. The PALPA calls for an alternative
way of thinking about aphasia.

There is little doubt that clinicians can meet this challenge. Educators and trainers,
both at undergraduate and postgraduate level, are already teaching psycholinguistic
techniques, and many therapists already apply the hypothesis-driven approach to
assessment and therapy (e.g. Jones and Byng 1989). Prior to the appearance of the
PALPA clinicians have employed considerable ingenuity in developing their own,
‘home-made’, tasks with which to test hypotheses (e.g. Jones 1989, Cook 1991).
Kay et al. (1992) have also done their best to equip clinicians for using these tests.
The Introduction provides an accessible theoretical background and each test is
accompanied by a brief description of what it explores, guidelines for its interpretation
and suggestions about where to go next. Their pointers to relevant references are also
extremely useful.

Even with these ‘safeguards’ the PALPA does contain a number of potential hazards.
One danger is over-testing. It seems churlish to criticize the comprehensiveness of the
collection, especially as the authors state repeatedly that the PALPA must be used
selectively. However, they also seem to believe that clinicians have limitless
assessment time. It is not atypical to find five or six further tests suggested in the ‘Where
to go next’ section. Also some potential corners have not been cut; for example, test
53 explores the naming, reading, repetition and spelling of 40 pictures. I cannot see
why a frequency ranking is not given for these items. This would enable clinicians to
get a rough idea about whether frequency is a factor in naming, without necessarily
having to administer the 60 additional pictures in test 54. If the PALPA is to be clinically
useful therapists must develop economical ways of using the tests. Towards this aim,
more guidance about how to use the collection so screen patients would have been
helpful. This could be helped again through reorganization; for example, tests
exploring the more esoteric areas of language processing, such as morphological skills,
could be siphoned off into a ‘Supplementary PALPA’, leaving the more mainstream
tests in the body of the collection.

Another problem, which is shared by all assessments, relates to the interpretation of
the test scores. It is a pity that not all the tests have normative data. For example
a third of the spelling tests lack control measures and, as I know, this is an area of
language processing that can be problematic for non-dysphasic people. However, the
problem is not overcome solely through comparisons with normal language users.
Dysphasic patients, for a variety of reasons, are likely to make more errors than
controls. At issue is how many errors constitute a deficit? This is not at all obvious given
that the tests vary a great deal in their inherent difficulty—for example some tap
tablet levels of processing while others explore relatively isolated skills.
Inconclusive test results may encourage clinicians to go on testing until they get more
emphatic data, thus furthering the danger of over-testing.

It is difficult to see a solution to this problem—rigid scoring criteria being clearly
inappropriate. One response would be to encourage the sharing of test data, so that
clinicians can compare their patients with other reported subjects. The authors mention
that piloting has been done with 25 dysphasic people, yet they do not make the results
available. This may be because they feel that acontextual group data are relatively
uninformative about these tests. However, the results could give clinicians a broad idea
of the range of dysphasic performance. Perhaps a better alternative would be to
encourage the creation of a ‘PALPA Case Book’. This would report test scores in the
context of analyses of single patients, thus providing practising clinicians with an
invaluable frame of reference. Another approach would be to place less emphasis on
test errors and more on success. When a patient passes a measure we can be reasonably
sure that the processes underlying the task are intact (providing our theoretical
assumptions are correct). Therefore, unlike failure, test success provides fairly
unambiguous evidence about the functioning of the language system. These results
can also be very useful for our patients, and not simply as a morale booster. Patients
may overestimate the extent of the damage to their language system. Success on a test
can alert them to the areas of processing which are still intact, and which might provide
them with resources to help with the other problems.

A third general concern is that the PALPA will discourage more ‘functional’
assessments of patients. This is not to say that the insights provided by the PALPA are
functionally irrelevant. For example, discovering that a patient reads via sublexical
routines has many practical implications. From this we can advise relatives and friends
about what types of word are likely to be easiest and can suggest strategies, such as
phonetic spellings, to make words more accessible. Yet the tests themselves often
explore skills which are of little direct interest to our patients—non-word reading
and letter naming being prime examples. They also fail to address patients’ specific
concerns, such as whether or not they will be able to read cherished books, the cricket
scores or product names in the supermarket. This may have more than just emotional
implications. The lack of functional relevance may prevent some patients from
displaying their full abilities. Kay et al. (1992) make no apology for the ‘abstract’ nature
of their collection, and state explicitly in their Introduction that alternative measures will be needed to explore language use. Whether it is possible to produce tests which combine functional relevance with the control of psycholinguistic parameters remains in doubt. Such measures would certainly be extremely welcome (Parr 1992).

Perhaps the greatest danger is the assumption that the PALPA will somehow provide clinicians with a direct lead into therapy. These tests undoubtedly give more direction than previous aphasia batteries, which offer little more than a broad description of the surface symptoms. They supply information about which language processes may be impaired and which are intact and this, in turn, can help the clinician identify a target for therapy. Yet, of itself, this level of hypothesizing does not constitute the theory of intervention (Byng et al. 1990). Such a theory must also take account of factors outside the strict cognitive assessment. These would include the patient’s communicative needs and personal interests, his/her language environment, his/her current strategies and problem-solving skills, the type of therapeutic interactions which seem to promote success, the motivation for therapy and ability to participate in different regimes, etc.

The specific relationship between the cognitive deficits identified by assessment and the content of therapy is also far from simple. The clinician must still decide whether to try to reinstate the damaged systems, regain access to blocked levels of processing or exploit residual skills through a compensatory therapy. Even if the deficit is the focus of therapy, the relationship between the treatment and its effect on the damaged language system is unclear (Caramazza 1989, Hillis 1993).

This gulf between the ‘diagnosis’ and therapy exists partly because the nature of the language model is still so poorly understood. This can be seen from the PALPA tests themselves. Take tests 47 and 48 as examples; these explore the patient’s ability to match a spoken or written word to a picture; the presence of semantic, visual and unrelated distractors. The semantic distractors hold a variety of relations to the targets. Some share category membership (e.g. apple and orange), others are related by function (e.g. ladder and steps) and some by real word/event connections (e.g. cobweb and spider). Poor performance on these tests, together with a tendency to select the semantic distractors, would suggest that there is a ‘semantic deficit’—a conclusion which would be supported by failure on other assessments within the section. However, the precise nature of this deficit remains a mystery. This is mainly because we do not understand the normal workings of the semantic system—and the diverse nature of the semantic associations tapped by the PALPA tests reflects this lack of understanding. It is highly probable that patients diagnosed as having a ‘semantic deficit’ in fact have a variety of problems. For example some may have impoverished semantic representations while in other patients the organization of the semantic system may have broken down—they may have lost notions of category membership or lack connections between related items. We simply know that all the patients seem unable to make the various types of meaning discriminations demanded by semantic tests. In other words a ‘semantic deficit’, as yet, can only be defined as a tendency to make errors in semantic assessments.

The diagnosis of a ‘semantic deficit’, therefore, does not solve the therapy problem. One reasonable treatment response would be to assume that if we can improve our patient’s performance on tasks similar to those used in the semantic tests, we may be improving the semantic system itself, and thereby achieve the desired speech and comprehension aims. This approach has been adopted by clinicians and seems vindicated by evidence of improvements post-therapy (e.g. Scott 1987). However, there are difficulties. Firstly there is the theoretical problem that we do not know exactly how (or if) the therapy has affected the workings of the semantic system. We are only assuming that progress has occurred because of the post-therapy gains. There is also the difficulty that amnic patients who do not make errors on semantic tests also seem to benefit from semantic tasks in therapy (Howard et al. 1985, Marshall et al. 1990). This suggests that the therapy may be having a general effect on the naming system, rather than specifically righting a semantic deficit.

Of course therapists do not simply target patients’ deficits in their treatment. They also capitalize on skills. As already suggested, one of the major advantages of the PALPA is its capacity to identify areas of intact processing, which can then be exploited in therapy. Jones (1989) describes a treatment programme which aimed to facilitate phonological access in naming by drawing upon the patient’s preserved semantic abilities, and one of the single cases in Marshall et al. (1990) capitalized on two preserved skills—the ability to access phonology directly from the written word and the ability to make fine semantic discriminations. Of course we still do not know precisely how these therapies worked. We can only conclude that it does seem helpful to encourage patients’ conscious use of residual skills.

To suggest that we do not fully understand the links between the results of a cognitive assessment and the content of therapy is not really a criticism of the PALPA or of the individual tests within it. It is simply a reflection of our current state of knowledge. If we are to develop this knowledge, building a body of theoretically driven treatment studies is essential. As a minimal requirement these should supply detailed cognitive analyses of the patient’s skills and deficits, explain the content and rationale of therapy, evaluate treatment effects and attempt to interpret them against the presumed cognitive model (e.g. Hillis 1993). By comparing these individual studies we may address some of the questions about which therapies are helpful for which problems, and begin to understand how therapy enhances or alters the damaged language-processing system.

Comparisons between studies may be difficult if patients have been investigated on different measures. Until now this has been a problem. The appearance of the PALPA should bring more order to the enterprise. Now both research and clinical studies can draw upon the same assessments, and direct comparisons between their data will become possible. By laying down this common ground the PALPA will hopefully further our theoretical understanding of aphasia and our knowledge about how to treat it.

References


PALPA: the proof of the pudding is in the eating

JANICE KAY, RUTH LESSER and MAX COLTHEART
Exeter University, Exeter, UK

Introduction

We want to begin by thanking the five authors who have provided commentaries on the PALPA tests for their detailed and insightful comments. Although we have not agreed with all of their comments (and, indeed, strongly disagreed in places!), we have been grateful for the chance to engage with them in debate, since we believe that this can only be to the good of assessment of dysphasia in general.

The comments that the commentators make naturally divide into a number of sections, to do with theory, methodology and practice. Some of the authors are concerned with all of these areas, others have focused on one in particular, but we will take their points as they arise in our arguments.

Theoretical bases of PALPA

The PALPA battery takes a cognitive neuropsychological approach to understanding acquired language disorders. At the beginning of the Introduction to PALPA we have included a simple box-and-arrow diagram which shows how abilities such as hearing and understanding a word can be divided into a number of different components. The diagram provides a framework, or functional architecture, with which to characterize aspects of language performance. In our experience this is a useful way of getting people to think about how particular language tasks (such as repeating words or reading aloud), traditionally used in speech-and-language clinics, may be carried out—but in our Introduction it is little more than that. The diagram itself is simply a description with some structure to it. In their commentary, Ferguson and Armstrong suggest that the whole cognitive neuropsychological approach is, in fact, fundamentally misdirected, and one of the things they condemn is the emphasis that is placed on the quest to derive a basic architecture of various language abilities. This approach, they suggest, has been criticized for 'failing to increase our knowledge of how language is stored, accessed and processed', and they cite Seidenberg's (1988) paper to bolster their claim. While Seidenberg is indeed critical of cognitive neuropsychology (though he styles himself as 'an outsider who is fundamentally sympathetic to the enterprise'), he does not dismiss it in the cursory way that Ferguson and Anderson do. Thus, he states, 'If we are now in a position to pursue more rigorous theories of language processing, it is in part because of the foundations laid by this research.' Seidenberg's brief is rather to bring home the point that models of the 'functional architecture' cannot advance much beyond a level of description unless fundamental questions concerning the nature of knowledge representations and processing mechanisms are addressed. Indeed, he argues that if the aim of deriving a functional architecture is effectively an end in itself, it may fatally disrupt further progress towards addressing these questions (in much the same way that classification by syndrome label—Broca's aphasia, Wernicke's aphasia—may have obscured attempts towards a deeper understanding of such disorders).

On the other hand, Seidenberg briefly discusses investigations from the same mould, which do seem to provide more than mere redescriptions of the observed findings. These studies have in common 'the right form'; that is, they are based on independently motivated theoretical grounds which allow key empirical predictions to be made and tested. He concludes by noting that, rather than dismissing the whole enterprise, 'we are simply going to have to bite the bullet and attempt to develop more explicit—yes, computational—models in order to answer them'. One of us has been doing precisely that. The functional architecture represented as a box-and-arrow diagram at the beginning of the Introduction to PALPA is the functional architecture of a computational model of visual word recognition, reading aloud and spelling, the DRC model (Coltheart et al. 1993, Coltheart and Rastle 1994). Since a computational model is by definition expressed in the form of a working computer program, the development of the DRC model has required that the ways in which the boxes and arrows of this functional architecture actually work be explicitly specified. Not only does this model successfully simulate a wide variety of phenomena observed in studies of normal readers, but it can also be artificially 'lesioned' so as to simulate various forms of acquired dyslexia (this is 'computational cognitive neuropsychology'). Detailed patterns of selective preservation and impairments of different aspects of reading, reported in published cases of surface and phonological dyslexia, have been successfully simulated (Coltheart 1995). Hence we are now in a position to do what Armstrong and Ferguson wish: we can go beyond just proposing a functional architecture to make explicit proposals concerning how (some aspects of) language are stored, accessed and processed.

As we stated above, and now emphasize, in the PALPA we have used a simple 'functional architecture' of word recognition and production as a guide, because we have found it to be a helpful way of introducing this way of assessing and thinking about language disorders, not because we believe it to be a particularly valid
Clinical Forum

'‘model’ of language structure. It is this way of thinking that we believe can provide useful insights into both diagnosis and therapy. To echo Wertz, we contend that, in this case, the proof of the pudding is in the eating. Let us use work on auditory comprehension disorders as an example.

In a recent review article, Blumstein (1994) suggests that nearly all aphasic patients show some problems in discriminating phonological contrasts, that such perceptual problems are more likely when test stimuli contrast by a single feature than when they contrast by two or more features, that the perception of place of articulation is particularly vulnerable, and that these problems emerge for the perception both of real words and nonsense syllables, though nonsense syllables are more prone to error (Baker et al. 1981, Blumstein et al. 1977, Jauhiainen and Nuutila 1977, Miceli et al. 1980). Blumstein (1994) also notes that non-brain-damaged people show a similar profile when perceiving speech under difficult listening conditions (Miller and Nicely 1955). On the basis of her review of a series of experimental studies, Blumstein claims that ‘aphasic patients do not have a deficit specific to the extraction of the spectral patterns corresponding to the phonetic categories of speech’, but that rather ‘their deficit seems to relate to the threshold of activation of the phonetic–phonological representation itself or to its ultimate contact with the lexicon’. In fact, Franklin (1989) shows how detailed investigation of individual patients can demonstrate a variety of different levels of breakdown of auditory processing, so that while one patient may have difficulties which can be assigned to a ‘lexical level’ of speech comprehension, another may have problems which appear to be located at a much earlier stage of speech analysis.

Differences can be revealed through characteristic profiles of performance on a number of tasks such as those discussed by Blumstein, and used in PALPA, which build in dimensions such as word–non-word, distance of phonological contrasts, voicing, manner and place of articulation, which Blumstein notes to be of importance, as well as other linguistic dimensions such as word frequency and imageability, which also have independent reasons for inclusion. Morris et al. (1996) use such tasks with an individual patient, J.S., to demonstrate that he has a specific problem in processing auditorily presented material which appears to occur at an ‘early stage of auditory analysis’, as well as a more central semantic deficit. J.S. showed a particular pattern of performance on auditory tasks: he found it difficult to make minimal pair contrasts, for example (although he was not at chance), he was better with ‘maximal pairs’ (pairs which differed by three features) than ‘minimal pairs’ (pairs which differed by only one feature), and he improved when instructed to lip-read. The authors therefore decided to focus on trying to improve his ability to carry out auditory single-word processing through a tailored therapy programme which systematically manipulated the number of distinctive features between two stimuli, and which encouraged lip-reading. They observed that a relatively short period of treatment resulted in significant improvements on phoneme discrimination tasks which generalized to test items that were not used in therapy. Repetition of words was also better (which was attributed to J.S.’s newly enhanced auditory perceptual abilities). His functional comprehension was also aided by his improved ability to make use of lip-reading (personal communication).

Now, it is possible to use the basic ‘functional architecture’ of auditory word comprehension that we illustrate in PALPA as a starting-point for a study like that of Morris et al., since it provides a logical structure to the use of relatively familiar phonological tasks such as minimal pair contrasts, and relatively new ones such as auditory lexical decision. A cost is that it may be taken rather rigidly as a proposal of the way in which auditory word comprehension is achieved. A description of the model implies, for example, that acoustic information is transformed into a more abstract phonemic (segmental) representation prior to lexical access, and that such information is processed in a strictly serial fashion (i.e., accessing the lexicon takes place after auditory phonemic analysis is completed). From the perspective of theoretical research it is important to realize that alternative theoretical proposals exist. Some would claim, for example, that such procedures, although still based on a phonemic analysis of the speech signal, do not operate serially, but in cascade (e.g., McClelland and Elman 1986). Others argue, more radically, that phonetic features are taken onto contextual–lexical–level forms in a distributed representational system (e.g., Marslen-Wilson 1993, Marslen-Wilson and Warren 1994). From the perspective of theoretically driven assessment, however, it is possible to use the basic framework to drive assessment and therapy, but not to be trapped within it. Thus, as well as the basic PALPA-like tasks, Morris et al. use a variety of other tests which are derived from psychophysical work on early auditory processing abilities of non-brain-damaged people (i.e., gap detection, formant frequency discrimination, frequency modulation detection and pitch discrimination). They use these tests to show that J.S. has difficulty with gap detection and frequency modulation, which have to do with fine-grain temporal processing, and that he is impaired not only in processing speech sounds, but also non-speech sounds (see also Best and Howard 1994, for a similar proposal concerning 'word sound deafness'). Such difficulties may account for his problems with minimal pair discrimination. As well as providing a theoretical account of J.S.'s difficulties, the authors also contribute to the issue of whether there is a normal 'speech mode' of perception that is involved only in processing speech and not in processing non-speech sounds. They suggest that 'if such a speech mode exists, impairment to it could give rise to speech-specific word sound deafness; if it does not, then word sound deafness will always be seen in the context of impairments to non-verbal auditory processing. We note that J.S. provides no support for the notion that a speech-specific mode of processing is selectively impaired in word sound deafness.'

It is our contention that the enterprise of cognitive neuropsychology has produced important theoretical insights into the nature of language processing (and will continue to do so), and that its methodology and techniques can drive assessment, management and therapeutic intervention in dysphasia. The basic framework adopted by PALPA reflects this, and although the model that it espouses for teaching purposes can foster an unnecessarily restrictive view of language, we have found that most practitioners are capable of using it in a flexible way (indeed, it would be patronizing to assume otherwise).

Ferguson and Armstrong criticize the emphasis placed on single-word processing throughout the PALPA tests. They suggest that this focus is a 'rather outdated view (as far as linguistics is concerned).’ There are a number of issues here. One concerns the use of ‘single-word’ tests. Another has to do with the authors' view that we do not take linguistic theory into account. Let us take our use of single-word tests first. This emphasis is common enough in investigating acquired language disorders, whether from a traditional aphasiological perspective or from that of cognitive neuropsychology. Such tests also constitute much of the bread-and-butter work of cognitive psychologists who are interested in issues concerning the structure and organization of the mental lexicon. While this is
clearly not sufficient justification, a more important reason is that it is misguided simply to dismiss the data—which seem to us to be of considerable theoretical and practical significance—that have been (and can be) gathered through use of such tasks. We shall give brief examples, each of which illustrates this point in a slightly different way. First is the recent work by Plaut and his colleagues, which demonstrates how powerful new tools such as connectionist networks can model, or implement, patterns of normal and acquired dyslexic performance on ‘single-word tasks’ such as reading words and novel letter strings (e.g. Plaut and Shallice 1994). Second, such tasks have been used to considerable effect in recent PET-scan studies of the neural representation of lexical processes (e.g. Howard et al. 1992). Third, much has been learned about general principles concerning the organization of lexical–semantic knowledge through detailed use of single-word and picture material in dysphasia and in progressive dysphasia (Snowden et al. 1989, Hodges et al. 1992). Arguably this work, also including representation of category-specific and modality-specific knowledge (Caramazza 1994, McCarthy and Warrington 1994), could not have progressed so rapidly by any other means. Finally, studies such as that carried out by Morris et al. (1995), described above, indicate how detailed analysis of single-word (or, rather, single-item) processing can have important implications not only for therapy but also for practice (assessment, management and therapy).

Throughout their commentary, Ferguson and Armstrong criticize us for our lack of sophistication in linguistic theory. By this they mean that our view of the lexicon is not informed by current accounts in linguistics (we must stress again that, for the purposes of teaching, our view of lexical processing illustrated in the Introduction is greatly oversimplified). But we would argue that as much has been learnt about the structure and organization of the ‘mental lexicon’ from cognitive studies as from linguistic theory. Ferguson and Armstrong also mean that the model we have used to illustrate single-word processing—our view of the lexicon, if you like—does not map onto a model of sentence processing in any clear way. Now, of course, there is an immense research effort directed towards the investigation of sentence processing and its disorders, with almost as many theoretical approaches as research groups (see Safran and Schwartz 1994). As authors of PALPA we do not favour one account over another, so we take a neutral position on how ‘sentence processing’ and the ‘lexicon’ relate. Furthermore, we think that whether ‘different levels of language functioning…work on differing linguistic principles’ is a matter of experimental inquiry, involving not only ‘language’ and ‘ discourse’, but also other mental faculties such as working memory and switching/attentional resources. We do not, however, think that this means that we cannot include sentence processing materials (see also Basso, this issue). One of the purposes of PALPA (though certainly not the primary purpose) is to be a research tool which provides a springboard for further investigation: it is not designed to test between competing theories. There are certain findings that characterize some forms of grammatical disturbance in dysphasia: that the passive voice of a sentence is more difficult to process than the active, for example, and that semantically reversible sentences pose more problems than non-reversible ones. Whether for the goal of theoretical research, or for assessment and treatment in the clinic, any useful sentence-processing task needs to build in known dimensions such as these. There is also a body of research which indicates that thematic roles cued by verbs which have converse relations (e.g. buy/sell) can be confused by some dysphasic individuals, as can constructions in which subject–object relations are not made explicit in the spoken (or written) form. We included sentences of these types in PALPA, not because we feel any need to subscribe to the psychological reality of linguistic theories such as Government and Binding Theory (the motivations of such theories arise from a totally different orientation), but for a different purpose. It is for their potential for discriminating between individuals whose processing disorders may be attributable to difficulties with word order as such (including possibly restrictions of attention and/or working memory) as detected through pragmatically reversible sentences, and individuals who appear to confuse thematic roles where word order cannot be the problem. The different implications for planning therapy for each case could be of some consequence.

Finally, in this section, we need to mention Ferguson and Armstrong’s claim that we ‘largely dismiss sociolinguistic views of language’, including discourse and discourse analysis. We certainly need to dispel this impression that we do not place a value on this work, especially as one of us is actively involved in investigating the functional adequacy of aphasic language from these perspectives and applying it in intervention (Lesser and Milroy 1993, Lesser and Algar 1995). In our brief Introduction to PALPA we simply have nothing to say about assessment of discourse exchange, text, etc., just as if we were writing a textbook of Government and Binding Theory, we would also have little to say about the study of sentences in discourse (though see Sperber and Wilson 1986, and KEMPSON 1986a, b, who have). We are, as Ferguson and Armstrong note, writing within the context of a particular approach to understanding acquired disorders of language, but not to the exclusion of other perspectives with which we believe it can happily coexist. Rather than focusing on what Ferguson and Armstrong perceive to be ‘inherent limitations’, we prefer the more empowering approach that considers PALPA as another, powerful, arrow in the aspahologist’s bow.

From theory to practice: how easy is PALPA to use?

As Marshall notes: ‘the PALPA asks the clinician to adopt an experimental approach to the assessment of patients. Some reasonably specific hypotheses must be formed about the nature of the patient’s language impairments before the PALPA is taken out of the cupboard. The clinician must decide which assessments to choose in order to test these hypotheses and understand how to interpret the data which they supply. Above all, the PALPA demands an ability to conflate results from different assessments in order to deduce the underlying functioning of the language system. There are 60 tasks in the battery, and although some of them are more useful across a range of patients than others, there will be many possible points of entry. Ferguson and Armstrong find this a potential problem, relying as it does on a clinician’s judgement, or nous. They claim that it presupposes that the clinician has a better way than PALPA of detecting impairments in the first place. However, this misses the point. We believe that clinicians do initially make

1We would be appalled if PALPA discouraged more ‘functional’ assessments of patients. As Marshall notes, however, ‘this is not to say that the insights provided by PALPA are functionally irrelevant’, and it is worth mentioning that several of the PALPA tests and materials can be presented in a more functional way; reading aloud non-words, for example, can be presented as a task of directing a route through different streets of a road map (e.g. Fulford Road, Chaplin Street).
hypotheses about the particular language difficulties that a patient is experiencing, which are drawn from a variety of sources (e.g. clinical interview, a standard test such as the BDAE, assessing performance in a functional setting, a patient’s or
partner’s own report). Since there are many potential points of entry to assessment
according to each presenting pattern, one of PALPA’s functions is to allow the
clinician full control and management by providing a logical and rigorous
framework with which to structure assessment. But where to begin must depend on
the assessor’s judgement. Although we do not claim that this is an easy skill to
acquire, we believe, like Marshall, that clinicians are equal to the task. As Marshall
notes, prior to the appearance of PALPA, clinicians have employed considerable
resourcefulness in developing their own, ‘home-made’, tasks with which to test
hypotheses (e.g. Jones 1990, Cook 1991)—though, as Basso comments, ‘this
obviously necessitates much time and ingenuity on the part of the examiner and is
not possible in a clinical setting’. PALPA provides well-controlled ready-made
tests, so that for some things clinicians will no longer have to compile their own
assessments. In our experience, clinicians will gradually accumulate a collected
‘case law’, indicating which tasks may be useful to begin assessment with which
‘type’ of patient.

Marshall picks up on another related point which is more difficult for us to
answer, at least at the moment. The PALPA collection may seem to be too
extensive, particularly if there is an unstated assumption that assessment time is
unlimited (not one that we hold, incidentally, and one that is difficult to sustain in
these hard days of ‘purchasers’, ‘providers’ and ‘outcome measures’ within the
British National Health Service). We counsel strongly against slavishly running
through a collection of tasks in a particular domain; this is completely counter to
our motivation in developing PALPA. It is nonetheless true that if PALPA is to
match its clinical promise, therapists must develop economical ways of using the
tests. It is also true that there are corners to be cut if one wants ‘quick and dirty’
information about the effects of particular factors on performance, rather than
experimental rigour. Marshall comes up with a number of excellent suggestions for
how to improve its use in the clinic, and we shall consider this issue again in the
section on future directions.

Psychometric standardization, validity, reliability and normative data

Wertz spends some time in discussing why psychometric standardization, and
related issues of validity and reliability, are important in clinical assessment. He
suggests that although lack of standardization of PALPA is not a fatal flaw, it does
remain work to be done. But is this the case? Some researchers have recently
questioned whether psychometric concepts such as validity and reliability should
be applied to certain ways of investigating cognitive impairment (e.g. Bates et al.
1991, Caramazza and Badecker 1989, McCloskey 1993). They have claimed that
cognitive neuropsychological assessment might require a set of data analysis
techniques different from, but complementary to, more traditional psychometric
measures (such as measures of dynamic change: Hoffman et al. 1992). Let us
call exactly why validity and reliability may not be meaningful in this context.

Issues of validity frequently arise in neuropsychological practice—for example, if
a particular cognitive test is proposed as a measure of the integrity of frontal lobe
showing a relationship between test performance and CT scan data from frontal
lobe-damaged persons). Such data are needed to show that the test does measure
what it purports to measure; that is, that the test is valid. But the situation is
different with PALPA. Consider its non-word reading subtest. What does this
subtest purport to measure? The answer is: non-word reading. So the question of
validity simply does not arise here (as it would if, for example, this non-word
reading subtest were proposed as a measure of the integrity of certain left temporal
lobe structures). This point is applicable to PALPA subtests in general. In this
sense, then, PALPA does not need validation.

Reliability

There are unique issues here too. Suppose one were to carry out a major test–retest
study, administering PALPA to a group of aphasics on two occasions. Correlations
between test and retest are one conventional measure of reliability. But this is
inappropriate here. Shallice (1987) drew a distinction between disorders of access
and disorders of representation in aphasic patients: errors may arise because of
failure to access information that is still present in the language processing system,
or because of loss of such information from the system. He offered, as one way of
distinguishing between these two possibilities, item-specific consistency of
performance. If an aphasic tested with the same items on different occasions gets
the same ones wrong every time, that suggests that their representations are no
longer present; if instead there is much inconsistency from occasion to occasion in
which items are responded to correctly and which incorrectly, that suggests that
the difficulty is that access to representations is faulty, succeeding on some
occasions and not on others. So low test–retest reliability here is not regarded an
indication that the test being used is psychometrically deficient.

To put this another way: suppose a patient is given the same set of pictures to
take on two different occasions, and there are frequent instances in which a picture
is correctly named on one occasion and not on another. It might seem natural to
explain this by saying that the test is unreliable, but that is not an explanation, it is
merely a redescription. An explanation would be an account of the patient’s
naming disorder that gave a reason why a picture named correctly on one occasion
can be named incorrectly on another (for example, an account of the nature of the
the access procedure which made that procedure fluctuate in efficiency across items or
occasions after damage to the brain). Matters are different when normal subjects are
being assessed: here there are no reasons intrinsic to the testees to expect low
test–retest reliability.

Wertz says ‘In the PALPA, we do not want to infer a patient has processing
difficulty in one module or performance is affected by one psycholinguistic variable
today and discover neither appear impaired tomorrow’. But why not? As Shallice
has argued, this pattern might be telling us something important about the nature of
the patient’s impairment (not something negative about the nature of the test
instrument).

Normative data

The use of norms arises because we are unable to do what we really would like to
do, and that is to have premorbid assessments for all our patients. Imagine that

Now PALPA data from someone who has recently become aphasic could be compared directly with data from that same person collected before he/she became aphasic. Each person would be his or her own control, and norms would never even be thought of. Since this 'ideal' situation does not exist, we might want to estimate how the patient would have performed premorbidly by referring to data from a normative group of non-aphasic subjects. Wertz points to some of the problems here, such as how to define what is normal. If the criterion for normality is, say, within 2 standard deviations from the mean of the non-aphasic sample, then a perhaps 2% of the members of any normal control group will be classified as aphasic. There are many other problems. If we are using norms to assess a person's spelling ability, for example, it surely does not make sense to use the same norms for a patient who was premorbidly a university lecturer in English and a patient who left school at the age of 15. We really need a separate normative group for every patient, matched to the patient on such variables as age and educational history, if that group is to offer a sensible estimate of the patient's premorbid ability. That is clearly out of the question. These difficulties apply to all aphasia assessment instruments—PALPA included. One partial solution to this problem that PALPA provides is that many of its subtests are so easy for non-aphasic people that normal accuracy is 100%. For example, even minimal literacy would provide perfect performance on letter-naming, letter orientation judgement, and cross-case matching. So if a patient's premorbid history is such that the tester can be confident that at least some degree of literacy was present, any errors on such tests are indicative of impairment. This is not, however, the case for all of the tests. Here, as Marshall cogently notes, what is at issue is 'how many errors constitute a deficit', and this is not simply answered by comparing it with normal performance. Marshall's answer is to encourage sharing of test data, so that profiles of performance can be compared between patients. But we would add a note of caution here: two patients might show closely similar performance on a particular task or tasks, and yet this pattern may correspond with a 'true' impairment for one but not for the other.

Where we depart from Marshall is in her suggestion that 'when a patient passes a measure, we can be reasonably sure that the processes underlying the task are intact.... Therefore, unlike failure, test success provides fairly unambiguous evidence about the functioning of the language system.' As non-brain-damaged people may function at ceiling on at least some of the tests (that is, one does not know at which point they would begin to fail), a more rigorous response measure (reaction times, for example, rather than accuracy) might in fact reveal differences.

Individual PALPA tests

The development of PALPA is ongoing. Informal feedback from clinicians, and from the commentators in this forum, show that included in our choice of 60 tests are some that appear to be of limited relevance, and some that do not do the job that they were designed to do. The tasks that appear to be most in need of revision, according to both our own informal 'market research' and to the commentators, are those to do with sentence processing. We still lack a commonly accepted explicit model of sentence comprehension and, given the range of difficulties which have been posited as underlying types of grammatical disorders, it would be premature to attempt to catch them all in one or two tasks. It is nonetheless clear that, for its implications for therapy, a test must include mapping of thematic roles which both do and do not involve processing of word order. The contrast incorporated in PALPA between adjective and verb predication has rarely, however, proved discriminating in our data from different patients, perhaps due to the limited number of sentences involved, and some revision of the sentence comprehension tasks is needed. Test 57, which is designed to test comprehension of individual verbs and adjectives used in the sentence—picture matching tasks, needs to be redesigned. As Marshall notes, it requires sentence-level skills in order to interpret the definitions. Nearly half of the items use synonym judgement, and a better procedure might be to ask for such judgements throughout. As well as modifications to the existing tasks, Basso notes, quite rightly, that there are some tests that could be added with advantage. She suggests, for example, that we should include a grammaticality judgement task that can assess ability to judge syntactic well-formedness (e.g. Linebarger et al. 1983). We intend to do this, using those grammatical violations to which 'asyntactic comprehenders' appear to be particularly sensitive, such as wh-movement (Linebarger 1990). Furthermore, given the postulated importance of the verb in sentence generation (Marshall 1994, Whitworth 1994), we should like to include further measures of verb processing. We are less enthusiastic about including an action-naming task as a measure of facility in retrieving verbs, as Basso suggests, given the variety of verb types beside those which concern actions, and the inherent distinction between verbs used as predicates and verbs used as names in isolation. Furthermore, we see less need to include an assessment of sentence production, as it is a relatively major undertaking in itself, and also since there are existing assessments that are available (e.g. Saffran et al. 1989, Menn and Obler 1990, Edwards et al. 1993).

We also see omissions elsewhere in the battery. Confrontation naming, which allows comparisons across a variety of 'output' tasks (e.g. reading aloud, repetition, written spelling, written naming) and examines whether there is an effect of word frequency on naming success (tests 53 and 54), does not allow one to observe whether syllable length plays a role (or whether there is an interaction) (see Nickels 1995). Similarly, comprehension and naming tasks will not pick up category-specific deficits (see McCarthy and Warrington 1994). Neither do they allow a thorough investigation of modality-specific disorders, although PALPA word—picture matching tests can be supplemented by the Pyramids and Palm Trees test (Howard and Patterson 1992).

We intend to incorporate such changes in a second edition of the battery. At the same time, though, we do not want to see an expansion of the tests to a level at which it becomes impractical. We therefore need to do some thinking about which tests to omit, or at least to distinguish as less likely to be commonly selected.

Marshall and Wertz pick out errors on individual items or tests. Some of these have already been spotted and amended in errata sheets for purchasers. Wertz also asks whether PALPA can be used elsewhere other than in the United Kingdom. We are pleased to tell him that an American version will shortly be available (following valuable suggestions for modifications from Berndt and her colleagues at the University Hospital School of Medicine, Maryland), that the original version is used in Australia, that it has also been adapted for use in Holland and Spain, and that other foreign language adaptations are planned.
PALPA’s use in therapy

PALPA is not designed to provide materials for therapy, although Marshall comments that in most tests the number of items is generous enough to provide pre- and post-therapy evaluation measures, both of treated functions and, if desired, untreated control tasks. Ferguson and Armstrong complain that the theoretical framework followed by PALPA does not provide guidance for whether therapy is to involve stimulation, relearning or compensation. (We note that the terms stimulation, relearning and compensation may not be theoretically useful in helping us to think about therapy, nor may they cast any light on the way(s) in which therapy actually turns out to work.) It is certainly true that, as Howard (1994) notes, ‘the advance in cognitive neuropsychology has been primarily in the identification and diagnosis of specific types of disorder...[but] an analysis of the level at which performance on a task is breaking down does not specify the way in which the disorder should be treated’ (Howard and Hatfield 1987, Howard and Patterson 1989, Caramazza 1989). A theory about the level of breakdown is not a theory of intervention (Byng et al. 1990). We agree with Howard (1994) that it may, however, place constraints on the set of possible treatment approaches. Take the therapy study carried out by Morris et al. (1996), and discussed above. The patient presented with extremely poor auditory comprehension. His ability to carry out ‘auditory analysis’—to judge, for example, whether simple sounds were the same or different—revealed that he had greater success with ‘maximal pairs’ (pairs which differed by more than three features), than ‘minimal pairs’ (which differed by a single feature). This led naturally on to a treatment programme which began with asking the patient to make simple judgements (with maximally different pairs), and then by increasing the difficulty of the judgement by using sounds that were progressively more similar to each other. Note, however, that the authors cannot say why this treatment had the positive effect it had, nor that this therapy is better than another at remediating the patient’s pattern of impairment; another may have worked just as well, or better. But this cannot be a reason not to carry on with this approach, nor to claim it is not useful. We agree with Marshall that ‘to suggest that we do not fully understand the links between the results of a cognitive assessment and the content of therapy is not really a criticism of the PALPA or of the individual tests within it. It is simply a reflection of our current state of knowledge’. Like Marshall, we urge readers to try as far as possible to publish the results of their studies (and not just the ones that ‘worked’), so that a body of theoretically driven treatment studies can be built up.

References


Hodges, J. R., Patterson, K. E., Oxbury, S. and Funnell, E. (1992) Semantic dementia:


