Making sense of progressive non-fluent aphasia: an analysis of conversational speech

Jonathan A. Knibb, Anna M. Woollams, John R. Hodges and Karalyn Patterson

The speech of patients with progressive non-fluent aphasia (PNFA) has often been described clinically, but these descriptions lack support from quantitative data. The clinical classification of the progressive aphasic syndromes is also debated. This study selected 15 patients with progressive aphasia on broad criteria, excluding only those with clear semantic dementia. It aimed to provide a detailed quantitative description of their conversational speech, along with cognitive testing and visual rating of structural brain imaging, and to examine which, if any features were consistently present throughout the group; as well as looking for sub-syndromic associations between these features. A consistent increase in grammatical and speech sound errors and a simplification of spoken syntax relative to age-matched controls were observed, though telegraphic speech was rare; slow speech was common but not universal. Almost all patients showed impairments in picture naming, syntactic comprehension and executive function. The degree to which speech was affected was independent of the severity of the other cognitive deficits. A partial dissociation was also observed between slow speech with simplified grammar on the one hand, and grammatical and speech sound errors on the other. Overlap between these sets of impairments was however, the rule rather than the exception, producing continuous variation within a single consistent syndrome. The distribution of atrophy was remarkably variable, with frontal, temporal and medial temporal areas affected, either symmetrically or asymmetrically. The study suggests that PNFA is a coherent, well-defined syndrome and that varieties such as logopaenic progressive aphasia and progressive apraxia of speech may be seen as points in a space of continuous variation within progressive non-fluent aphasia.

Keywords: progressive aphasia; frontotemporal dementia; language; cortical atrophy; neuropsychology

Abbreviations: FTLD = fronto-temporal lobar degeneration; LPA = logopaenic progressive aphasia; MDE = metadata extraction; MRI = magnetic resonance imaging; PNFA = progressive non-fluent aphasia; SD = semantic dementia; TROG = Test for the Reception of Grammar; VOSP-PD = Visual Object and Space Perception battery, Position Discrimination subtest; WCST = Wisconsin Card Sort Test; WPMT = word-picture matching test
Introduction

When Virginia McKenna (the wildlife campaigner and lead actress in the film Born Free) was awarded an OBE in 2004, a BBC radio presenter prone to speech errors announced that she had been honoured for ‘services to conversation’. The topic of this article is progressive aphasia, a condition whose hallmark is a gradual but relentless deterioration of the ability to converse. The aim of the study was to use detailed and quantitative analyses of these patients’ conversational speech to advance our knowledge of the component abilities which, when not conserved, lead to its decline.

The literature on questions regarding the spectrum, nosology, nomenclature, relationship to brain pathology, etc. of the progressive aphasia is extensive and growing (Neary et al., 1998; Mesulam, 2001; Kertesz et al., 2003; Gorno-Tempini et al., 2004; Kertesz et al., 2005; Josephs et al., 2006; Knibb et al., 2006). Although this literature will not be reviewed in detail here, one aspect of classification needs to be specified: the type of patient under investigation. The current study included a group of patients on the basis that their chief presenting complaint was a progressive deterioration of verbal communication. Our criterion for recruiting patients into this study excluded cases with a clinical diagnosis of semantic dementia (SD), defined according to the cognitive criteria consistently used in Cambridge (Hodges et al., 1992; Adlam et al., 2006; Hodges and Patterson, 2007). Despite the prominence of language disturbance in SD, it can be argued that this condition is not a form of ‘primary’ progressive aphasia (Mesulam et al., 2003), because the aphasia is just one reflection of an amodal semantic deficit (Adlam et al., 2006; Hodges and Patterson, 2007; Lambon Ralph and Patterson, 2008). The majority of patients studied would probably be described as suffering from progressive non-fluent aphasia (PNFA) by most researchers; but to those who split non-SD progressive aphasia into more than one clinical group, our sample might include a few cases who would be labelled with one or other of the following forms: logopaenic (Gorno-Tempini et al., 2004), adynamic (Esmonde et al., 1996), aphemic (speech apraxic) (Kertesz et al., 2003) or mixed (Grossman and Ash, 2004).

The characteristics of spontaneous speech in typical PNFA include effortful speech production, phonological and grammatical errors, and word retrieval difficulties. These characteristics are well-documented on a clinical level (Hodges and Patterson, 1996; Neary et al., 1998; Kertesz et al., 2003; Mesulam et al., 2003; Grossman and Ash, 2004; Ogar et al., 2007), and the various prototypes given in these references agree to a large extent. Other aspects of patients’ speech, however, remain under debate. For example, distortion of individual phonemes, and other features signifying impairment of the motor control of speech, are only inconsistently mentioned; and when these have been actively sought, they are reported in only a proportion of patients (Josephs et al., 2006). Also, many authors use the term ‘agrammatism’, but its significance is not always clear. Sometimes it describes a syndrome seen in Broca’s aphasia, with few complete sentences and a paucity of closed-class words and bound morphemes (pronouns, articles, prepositions, inflections, etc.). On other occasions, it has been used in a broader sense to refer to any derangement of spoken syntax. Some researchers report PNFA patients who fit the former description (Thompson et al., 1997), whereas others claim fundamental differences between syntax in PNFA versus Broca’s aphasia (Patterson et al., 2006).

Logopaenic progressive aphasia (LPA) has been described in a separate group of progressive aphasic patients, to be considered alongside PNFA and SD. Strong evidence has been adduced to document this pattern, which consists of slow speech, simple but correct grammar, frequent word-finding pauses, impaired repetition of sentences but not of single words, poor picture naming and moderate difficulty on sentence comprehension tasks, but good single-word comprehension and semantic abilities; it may be associated with impairment of phonological short-term memory (Gorno-Tempini et al., 2004, 2008). However, the question of whether this pattern is actually categorically distinct from PNFA, or whether on the other hand it represents one point in a space of continuous variation within PNFA, has not been answered. The present study is well-suited to addressing this question.

The core of this study is a quantitative analysis of speech in progressive aphasia. This combination of topic plus technique has a small literature—only four publications that we could find. The fact that there are so few might seem unsurprising to researchers who appreciate the time and effort required to elicit, record, transcribe and analyse conversational speech by aphasic patients. On the other hand, difficulty in conversing is—by the patients’ and carers’ reports and by clinicians’ judgement—the primary complaint in progressive aphasia, and to avoid addressing this is surely to miss an important point. Also, unsurprisingly, the existing reports do have limitations. Three examined only small numbers of patients (Thompson et al., 1997; Orange et al., 1998; Rogers and Alarcon, 1998). The fourth (Graham et al., 2004), with 14 cases, used a picture description task yielding speech samples of limited size and scope: the ‘Cookie Theft’ picture, though it has been popular and useful, rarely elicits naturalistically varied speech and almost never elicits complex syntax or morphology (Patterson and MacDonald, 2006). No study has yet examined a group of more than five patients in spontaneous, conversational speech, eliciting samples of sufficient length to make estimates accurate enough for quantitative analysis.

Further questions of interest that are addressable with a larger sample of patients include the relationship of quantified characteristics of conversational speech to both the results of neuropsychological testing, and to the distribution of brain atrophy. Patients with PNFA have cognitive impairments that are, by definition, mostly in the realm of language processing (primarily naming, comprehension of complex syntax and memory for verbal material); however, executive deficits have also been described (Broussolle et al., 1996; Nestor et al., 2003; Patterson et al., 2006) and it is not clear how widespread these are. Brain atrophy is usually present on magnetic resonance imaging (MRI) at the time of first presentation, the classical pattern being bilateral frontal and temporal atrophy, which is more severe in the dominant hemisphere. However, even this rather generic description fails to
apply in a significant proportion of cases (Sonty et al., 2003; Gorno-Tempini et al., 2004, 2006; McMillan et al., 2004; Josephs et al., 2006); (see Nestor et al., 2003, p. 2407, for a review of earlier studies).

To summarize, the aims of this study are as follows: first, to characterize conversational speech in a large group of patients with progressive aphasia; second, to establish which, if any, features are consistent across the group in terms of speech, cognitive testing and regional brain atrophy; and finally, to look for associations between these features that may help to make sense of the variability within the group.

Methods

Subjects

The following criteria were used to select patients for the study.

(I) Throughout the course of the illness, a gradual-onset, progressive disturbance of communicative function was the only or predominant (i) complaint; (ii) feature on clinical assessment; and (iii) factor contributing to impairment in activities of daily living.

(II) The neurological diagnosis at the time of testing was progressive aphasia only, and not another disorder of which it is a recognized feature (e.g. SD, corticobasal degeneration, motor neurone disease). The diagnosis of SD was based on the cognitive criteria consistently used in Cambridge (Hodges et al., 1992; Adlam et al., 2006; Hodges and Patterson, 2007). In brief, all had fluent speech with preservation of articulation, phonology and syntax. Anomia and impaired word comprehension were consistent features that contrasted with their preserved single-word repetition. All showed deficits on tests of non-verbal conceptual knowledge, such as the Pyramids and Palm Trees Test (Howard and Patterson, 1992). Although not part of the criteria, all showed characteristic anterior temporal lobe atrophy which was typically asymmetric (left being greater than right).

(III) MRI brain imaging showed no relevant abnormality other than atrophy.

Patients were invited to take part if they met all of the above criteria and were under review in the Cambridge Memory Clinic during the testing period (November 2005 to July 2006). Informed consent was obtained from all participants according to the Declaration of Helsinki, and the research programme was approved by the Addenbrooke’s Hospital Local Research Ethics committee.

Fifteen patients completed testing. All but one of the patients also fulfilled Mesulam’s criteria for primary progressive aphasia (Mesulam, 2001); one subject had mild symptomatic impairment of episodic memory and difficulty with complex motor tasks that was sufficient to affect her everyday activities; though this was much less disabling than her communication impairment. One other patient was tested <2 years from onset but fulfilled the criteria at follow-up. All patients were right-handed monolingual native speakers of British English, and none had any pre-morbid language disturbance. As part of a separate study, the patients were rated for the presence of motor speech impairment (dysarthria, speech apraxia or both), by consensus between three independent expert observers. Such impairment was found in 7 of the 15 patients.

Healthy control subjects were recruited from the volunteer panel of the MRC Cognition and Brain Sciences Unit. Subjects were chosen to match the patients in number and age (control mean age 71.5, patients 70.7; Cohen’s d = 0.09; Table 1).

Conversational speech

Elicitation and transcription

Each subject participated in a semi-structured conversation of 15–20 min with one of the authors (J.A.K.). The conversation was initiated by the interviewer, who suggested a broadly similar range of topics in each case; namely the subject’s employment, family, garden and other leisure activities; but did not prohibit other topics if they arose. The subject was informed of the purpose of the interview and was encouraged to ‘do most of the talking’. The interviewer took a mainly passive role in the conversation, and in particular avoided completing the subject’s utterances where possible; but otherwise tried to make the discourse environment as natural as possible.

Each conversation was conducted in a single session without breaks. No notes were taken during the conversation; with the patient’s prior consent, an unobtrusive digital voice recorder (Olympus DM-20) was used, which provided recordings of high quality, sufficient to detect subtle articulatory errors if present. Each recording was transcribed on a desktop computer within the following few days by the interviewer, using Adobe Audition 1.5 (Adobe Systems Inc.) for playback. Phonological errors and uninterpretable phoneme strings were given broad phonemic transcriptions. The transcript was annotated with reference to the Simple metadata extraction (SimpleMDE) guidelines, which form part of the Effective, Affordable, Reusable Speech-to-Text (EARS) project of the Linguistic Data Consortium of the University of Pennsylvania (version 6.2, June 2006). A summary of the method follows. The detailed guidelines can be downloaded from http://projects.ldc.upenn.edu/MDE/.

Discrete units

Using the SimpleMDE guidelines, the transcript was exhaustively segmented into syntactic units (so-called ‘SUs’) on syntactic criteria, although taking advantage of prosodic cues to syntax. These were categorized for the purposes of the present study as follows:

- complete: main clauses containing a subject and a finite verb;
- subordinate: subordinate clauses and relative clauses;
- coordinate: finite coordinated clauses lacking an explicit subject;
- modifier: an utterance consisting of a non-clausal constituent, which substitutes for or modifies a constituent in an immediately adjacent syntactic unit (including answers to interrogative ‘wh’-word questions);
- elliptical: an utterance consisting of a non-clausal constituent which is not a modifier;
- interrupted: a syntactic unit that remained unfinished because of interruption by another person;
- incomplete: a syntactic unit that remained unfinished for any other reason (including self-interruptions);
- empty: a syntactic unit consisting only of discourse markers (see next section); and
- uninterpretable: a segment of speech, bounded by two interpretable syntactic units or turn boundaries, of which the syntax is unclear.
Additionally, any syntactically incorrect syntactic unit was marked up as such (this relates to phrasal syntax—morphological errors were annotated on the individual word).

An illustrative example of an annotated transcription is given in Appendix 1.

**Word-level annotation**

The transcript was divided into words on standard orthographic criteria. Each word in the transcript was then annotated for word class (noun, verb, other open-class word, auxiliary verb, contracted auxiliary verb, pronoun, discourse marker, other closed-class word). The word types classified as ‘other closed-class’ were articles, conjunctions, demonstratives, possessives, prepositions and quantifiers (but not numerals). For each verb, the number of explicit arguments was also recorded, and any inflected word was annotated with its uninflected form. Each word-level error was coded according to whether the output was a real word, and whether it was related to the target by sound, meaning, both or neither (‘neither’ included errors with no obvious target). An uninterpretable phoneme string was counted as a single word.

Speech sound errors may arise from defects in the phonological representation of a word or in its articulatory planning, and this is sometimes evident within the error itself. Either or both may occur in PNFA. Phoneme-level errors have been considered more typical (Neary *et al.*, 1998), but dysarthria (which causes articulatory errors) and apraxia of speech (which causes both types of error) are both common in PNFA patients (Duffy, 2006; Josephs *et al.*, 2006).

Given a high level of expertise and specially designed articulatory tests, dysarthria, speech apraxia and aphasia may all be diagnosed in the same patient. However, it is not possible to say of an individual speech error, which of these three it arises from: a distortion may be either dysarthric or apraxic, and a substitution may be either phonemic or a gross articulatory error. Since in this study we were concerned

### Table 1 Results of speech analysis and cognitive testing

<table>
<thead>
<tr>
<th>Variable</th>
<th>Patient</th>
<th></th>
<th>Control</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean*</td>
<td>Range</td>
<td>Mean</td>
<td>Range</td>
<td>P</td>
</tr>
<tr>
<td><strong>General</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>–</td>
<td>6F, 9M</td>
<td>–</td>
<td>7F, 8M</td>
<td>–</td>
</tr>
<tr>
<td>Age-at-onset (years)</td>
<td>67</td>
<td>56.9–81.7</td>
<td>–</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Age at testing (years)</td>
<td>70.7</td>
<td>60.3–84.3</td>
<td>71.5</td>
<td>58.6–88.4</td>
<td>&gt;0.1*</td>
</tr>
<tr>
<td>Duration of illness (years)</td>
<td>3.7</td>
<td>1.6–6.9</td>
<td>–</td>
<td></td>
<td>–</td>
</tr>
<tr>
<td>Words analysed</td>
<td>1001</td>
<td>130–1889</td>
<td>1522</td>
<td>1088–2640</td>
<td>–</td>
</tr>
<tr>
<td><strong>Speech</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech rate (words/min)</td>
<td>93.2</td>
<td>29.6–145.5</td>
<td>153.9</td>
<td>113.4–202.3</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Phrase length (words/SU)</td>
<td>3.7</td>
<td>1.3–4.8</td>
<td>5.2</td>
<td>4.2–6.2</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Subordinate clauses (per SU)</td>
<td>4.9%</td>
<td>0–17.9%</td>
<td>20.7%</td>
<td>12–31%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Passive constructions (per SU)</td>
<td>0.4%</td>
<td>0–1–1%</td>
<td>1.9%</td>
<td>0–5.3%</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Embedded constructions (per SU)</td>
<td>0.0%</td>
<td>0–1–6%</td>
<td>1.1%</td>
<td>0–2.3%</td>
<td>0.0002**</td>
</tr>
<tr>
<td>Verb arguments (per verb token)</td>
<td>1.39</td>
<td>0.89–1.58</td>
<td>1.45</td>
<td>1.33–1.60</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>Inflected verbs (per verb token)</td>
<td>63.1%</td>
<td>33–80%</td>
<td>67.4%</td>
<td>49–81%</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>Contrasted forms (per word)</td>
<td>5.4%</td>
<td>0–8–9%</td>
<td>4.5%</td>
<td>2.7–7%</td>
<td>&gt;0.1</td>
</tr>
<tr>
<td>Modifier phrases (per SU)</td>
<td>12.1%</td>
<td>1–39–2%</td>
<td>12.8%</td>
<td>6–23–8%</td>
<td>0.41</td>
</tr>
<tr>
<td>Elliptical phrases (per SU)</td>
<td>6.8%</td>
<td>1–42–2%</td>
<td>2.8%</td>
<td>0–1–7%</td>
<td>0.01</td>
</tr>
<tr>
<td>Open-class words (per word)</td>
<td>41.9%</td>
<td>33–60–7%</td>
<td>45.9%</td>
<td>44–50%</td>
<td>0.042*</td>
</tr>
<tr>
<td>Noun–verb ratiod</td>
<td>0.74</td>
<td>0.27–2.71</td>
<td>0.97</td>
<td>0.69–1.36</td>
<td>&gt;0.1*</td>
</tr>
<tr>
<td>Speech sound errors (per word)b</td>
<td>4.6%</td>
<td>0–3–28%</td>
<td>0.4%</td>
<td>0–2–1%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Grammatical errors (per SU)c</td>
<td>17.6%</td>
<td>6–46–3%</td>
<td>3.2%</td>
<td>1–5–2%</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Closed-class word errors (per closed-class word)</td>
<td>0.38%</td>
<td>0–2–8%</td>
<td>0.13%</td>
<td>0–0–5%</td>
<td>0.004**</td>
</tr>
<tr>
<td>Semantic errors (per open-class word)</td>
<td>0.55%</td>
<td>0–1–4%</td>
<td>0%</td>
<td>0–0–3%</td>
<td>0.0005**</td>
</tr>
<tr>
<td><strong>Cognitive testing</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin card sort test (max 6)</td>
<td>2c</td>
<td>0–6</td>
<td>6c</td>
<td>3–6</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Graded naming (max 30)</td>
<td>9.2</td>
<td>0–23</td>
<td>24.1</td>
<td>20–27</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>TROG (max 28, chance 7)</td>
<td>18c</td>
<td>7–28</td>
<td>28c</td>
<td>25–28</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>Camel and Cactus Test (max 64, chance 16)</td>
<td>49.9c</td>
<td>35–59</td>
<td>59.0c</td>
<td>51–62c</td>
<td>0.0004</td>
</tr>
<tr>
<td>Word-picture matching (max 64, chance 6)</td>
<td>62c</td>
<td>57–64</td>
<td>64c</td>
<td>63–64c</td>
<td>&lt;0.0001**</td>
</tr>
<tr>
<td>VOSP position discrimination (max 20, chance 10)</td>
<td>19c</td>
<td>14–20</td>
<td>20c</td>
<td>17–20</td>
<td>0.049**</td>
</tr>
</tbody>
</table>

One-tailed t-tests are used except where indicated.

- For transformed variables, the back-transformed mean is given.
- Logit-transformed before analysis.
- Median quoted instead of mean.
- Log-transformed before analysis.
- Derived from data outside this study.
- **One-tailed Mann–Whitney U-test instead of one-tailed t-test.
- *Two-tailed instead of one-tailed t-test.

Additionally, any syntactically incorrect syntactic unit was marked up as such (this relates to phrasal syntax—morphological errors were annotated on the individual word).

An illustrative example of an annotated transcription is given in Appendix 1.
with the categorization and counting of individual errors, we did not attempt to make this distinction, but simply recorded whether the sound produced was closer to the target phoneme or to a different phoneme. When reporting errors, we therefore refer to them as 'speech sound errors', whether they occurred in spontaneous speech or in tests of confrontation naming.

The variables

The following quantitative measures were calculated (see Table 1 for a reference list of the variables):

Speech rate—to correspond as closely as possible to the clinical impression of ‘fluency’ in speech, all words were counted, including errors and words within self-corrections and repetitions, but not non-lexical output (such as ‘er’). The number of words in the transcript was divided by the total time during which the subject was ‘holding the floor’ (as opposed to producing supportive output such as ‘uh-huh’, ‘I see’, etc.).

Phrase length—the length in words of each discrete unit (SU) in a subject’s transcript was counted, after deleting repetitions and self-corrections. Interrupted, empty and unintelligible SUs were excluded, but incomplete SUs and those with syntactic errors were included. The distribution of these lengths is highly skewed, and the geometric mean was therefore used as the measure of average SU length.

Syntactic complexity—the frequency of subordinate SUs was calculated as a measure of the grammatical complexity of speech. Passive constructions were counted, as well as embedded clauses (in which a subordinate SU occurs within another SU). The mean number of arguments per verb across all tokens in a transcript was taken as a further measure of syntactic complexity. The proportion of verb tokens that were inflected, and the proportion of contracted forms among all words, were also calculated.

Elliptical phrase frequency—the proportion of modifier and elliptical SUs among all the SUs were calculated for each subject.

Word class ratios—the noun/verb ratio was calculated for each subject, using the total number of tokens of each class in the entire transcript. The proportion of open-class words was similarly calculated, using the sum of noun tokens, verb tokens and other open-class tokens divided by the total number of words marked up for class.

Error rates—speech sound errors were defined inclusively as all incorrect real words or non-words related by sound to the target, plus all unintelligible phoneme strings. Real-word errors related to the target by meaning were divided into closed-class word substitution errors and semantic errors (involving open-class words). The rate of each of these in a subject’s transcript was calculated as the number of errors divided by the total number of words analysed.

The category of grammatically unacceptable SUs was defined as the sum of unintelligible, incomplete and elliptical SUs, plus any other SU containing a syntactic error. The frequency of these was calculated for each transcript.

Reliability

To check inter-observer variability in the evaluation of the variables, a second observer re-transcribed three (20%) of the patients’ conversational speech recordings and analysed them for errors. The segmentation of the transcript into SUs and conversational turns, as well as the mark-up for word class, were done according to strict criteria as explained above and were therefore considered reliable by definition. For each of the variables listed in Table 1, the second observer’s score differed from the first observer’s by <12%. The only exceptions to this were the speech sound and grammatical error rates, where disagreement concerned not the output itself but its acceptability. A strict criterion was therefore chosen; SUs were judged according to whether they would be considered correct had they been written (given some flexibility for dialect and conversational style), while words were judged according to whether the produced phoneme sequence matched the target. Such strict scoring may seem ungenerous to the patients, in classifying as errors some utterances that would be treated as acceptable speech by more lax criteria. Where real connected speech is concerned, however, correctness is always a matter of judgement; and of course the same strict scoring was applied to the control transcripts.

Neuropsychological testing

A short battery of cognitive tests was administered to each patient, and most were also completed by the controls, except those for which data were available from previous studies by our research group. The tests included the Graded Naming test; the Test for Reception of Grammar (TROG); the 64-item Camel and Cactus Test of non-verbal semantic ability, in which the subject must choose which of four pictures ‘goes with’ a target picture (similar to the Pyramids and Palm Trees test (Howard and Patterson, 1992)); the 64-item Cambridge spoken word-picture matching test (WPMT) with 10 line drawings per item (Bozeat et al., 2000); the Wisconsin Card Sort test (WCST); and the position discrimination subtest from the Visual Object and Space Perception battery (VOSP-PD).

Due to time constraints, the TROG was reduced from 20 blocks of four items each to 7 blocks, for a total possible score of 28 and a chance level of 7. The seven most representative blocks were chosen by examining the item-total correlations among a separate group of PNFA patients tested previously, that did not overlap with the present group (T. Bak, unpublished data), in which the 7-block score correlated at r = 0.97 with the 20-block score (r² = 94%).

Brain atrophy

Regional brain atrophy was assessed by applying a visual rating scale to the patients’ MRI scans. This scale was developed for use in neurodegenerative disease, specifically to examine regions relevant to fronto-temporal lobar degeneration (FTLD), and has been validated against voxel-based morphometry in this patient group (Davies et al., 2009).

Full details of the method have been published elsewhere (Davies et al., 2009). In brief, coronal T1-weighted MRI scans of 15 brain regions in each hemisphere were inspected visually, and a rating of 0, 1, 2, 3 or 4 was assigned to each region according to the severity of atrophy, with reference to a set of pre-rated images. Each region was rated on the basis of a single MRI slice, four slices per hemisphere being necessary to cover all the regions.

The regions are as follows: orbitofrontal cortex, lateral frontal cortex (including dorsolateral and ventrolateral areas), anterior cingulate cortex, basal ganglia, temporal pole; anterior hippocampus, anterior parahippocampal gyrus, collateral sulcus, anterior fusiform gyrus, lateral temporal cortex (including superior, middle and inferior temporal gyrri), insula; mid-hippocampus, posterior superior temporal gyrus (including the planum temporale); and posterior hippocampus, posterior temporal cortex (including Brodmann area 37).

All ratings were carried out by the first author. Initially, 10 scans (a mixture of FTLD patients and healthy controls) were rated three times each to assess intra-rater reliability. Only areas which were rated identically on all three occasions in at least half of the scans were retained.
for the final analysis. Three areas were discarded for this reason, namely anterior cingulate, basal ganglia and orbitofrontal cortex.

Thirteen of the 15 patients in this study had MRI scans suitable for rating; two were unable to tolerate the scan procedure. Ten scans of age-matched healthy volunteers were added to these 13 scans. The rater was blind to the identity of the subject, to whether the scan came from a patient or a control subject, and to whether any pair of slices came from the same subject. The files were saved as bitmaps at 8-bit greyscale and 256 × 256 pixels in size, and were viewed on the LCD screen of a Dell Latitude laptop (screen resolution 1280 × 800 pixels) at an actual size of roughly 10 cm × 10 cm.

Results

Conversational speech

An average of 1000 words were transcribed and analysed for each patient, and around 1500 for each control participant. The data are summarized in Table 1 and Fig. 1.

Speech rate and phrase length

Speech rate (in words per minute) was significantly lower in the patients as a group, by roughly 40%; but three patients fell within 2 SDs of the control mean, confirming the clinical impression that slow speech is common but not universal in this group. The patients were spread fairly evenly, except for two outliers with the lowest values (Patients H and K). A similar pattern was found for phrase (SU) length, which was reduced in the patients by roughly 30% overall, but six patients fell within the normal range here, and again Patients H and K appeared as outliers with particularly low values.

Syntactic complexity

Occurrence of subordinate SUs was reliably reduced across the patient group: this class of utterance centred around 20% of all SUs for controls, but only ~5% for patients. Thirteen of the 15 patients fell outside the normal range on this measure. Both passive and embedded constructions were similarly reduced in frequency—unlike controls, most patients failed to produce even one example of either construction.

The verb-structure measures did not reveal any significant impairment at the group level. Both controls and patients produced a mean of approximately 1.4 arguments per verb, and inflected around two-thirds of verb tokens. Similarly, most patients fell within the control range on the proportion of contracted words.

The patients produced more apparently unjustified elliptical phrases than the controls, about 2.5 times as many on average; but there was considerable overlap between groups, such that only two patients (again Patients H and K) fell outside the control range, and the comparison only achieved borderline significance ($P=0.01$). Modifier phrase frequency did not differ between the groups.

Word class

The patients’ proportion of open-class word tokens was slightly reduced overall: the ranges were 44–50% in controls and 34–48% in patients. Although eight of the patients fell below the control

Figure 1 Results of speech analysis. Each alphabetical symbol represents the value of a particular variable for one of the patients, expressed as Z-scores with respect to the control results. Each patient is represented by the same letter in each column. The grey bar indicates the estimated range within which 95% of controls fall ($−1.96 < Z < 1.96$). Patient K produced no subordinate phrases, and therefore has no value for that column; this is indicated by a box around the symbol.
range, the two groups did not differ significantly considering the
number of variables examined ($P = 0.042$). Similar comments apply
to the ratio of nouns to verbs: while eight patients produced
fewer nouns per verb than any control, this did not achieve
statistical significance.

Error rates
The patients as a group made roughly 10 times as many speech
sound errors when compared with the controls, with one word in
23 typically affected as compared with one in 250. This was a
reliable finding: 14 out of 15 patients made more errors than any
control. Grammatical errors were also increased, with errors in
18% of SUs on average (compared with 3% in controls),
every patient fell outside the control range on this measure.
The rates of closed-class word substitution errors and semantic
errors were increased in the patients to a statistically significant
degree, but probably not a meaningful one, as the rates of these
errors were negligible in both groups.

Neuropsychological testing

Major group differences
As shown in Table 1, the patients performed significantly
worse than controls on all tests bar the VOSP-PD. Nonetheless,
there was considerable variability amongst the patients on
most of the tests, as is apparent in Fig. 2, with severe impairments
seen on only three tests: Graded Naming, the TROG and
the WCST.

Severe impairment on the WCST was almost universal, with
12 out of 15 patients completing two categories or fewer, whilst
most controls scored at ceiling (six categories). Three patients were
unable to complete even a single category. On the Graded
Naming test, only two patients were within the control range
(above 20 out of 30), and many of the patients had very low
scores, reflecting a marked anomia. Out of 312 naming errors
across the patient group, a large majority (284 = 91%) were
no-response or ‘don’t know’ errors; the others were mainly non-
word errors related by sound to the target (examples include
‘yakshak’ for ‘yashmak’ or ‘tamplee’ for ‘trampoline’). The
patients also showed considerable difficulty with syntactic com-
prehension as measured by the TROG. While most controls
scored at ceiling on this test, and none made more than three
errors, 12 of the 15 patients scored below this range.

Other tests
The ability to correctly identify semantic associates as measured by
the Camel and Cactus Test showed a reliable mean difference
between patients and controls, but the magnitude of this
impairment was no more than moderate, with a median score
of 86%. In the WPMT, almost all controls scored at ceiling, and
therefore a few very modest single-word comprehension impairments among the patients caused a statistically significant
group difference; despite all patients scoring >89%. On the
whole, abnormality on these receptive semantic tests was mild
and nothing like the degree of impairment typically seen in SD.
On the VOSP-PD test, all but two patients scored within the control range.

**Speech and cognitive tests: associations**

Six of the speech variables, and all the cognitive tests except for VOSP-PD, were selected for further analysis on the basis of a significant contrast between patients and controls, and of sufficient variance among the patients. The significant correlations between these 11 variables are listed in Table 2.

Speech rate, phrase length and elliptical phrase frequency were strongly mutually associated, with Pearson $r = 0.79–0.89$ surviving Bonferroni correction for the 66 pair-wise comparisons between the 11 variables. The correlation between speech sound and grammatical error rates also met the Bonferroni criterion ($r = 0.85$), and these two groups were linked by an association between speech sound errors and elliptical phrases ($r = 0.85$).

Strong, though less significant, associations were also found between the TROG, the Camel and Cactus Test and the WPMT; between the TROG and Graded Naming and also between the Camel and Cactus Test and the WCST (0.00076 < $P < 0.05$; $r$ or Kendall’s $τ > 0.53$). There was no correlation of this strength between any of the speech variables and any of the cognitive test results.

**Comparison with logopaenic and apraxic sub-syndromes**

A form of progressive aphasia named ‘LPA’ has recently been suggested as categorically distinct from both PNFA and SD. The characteristics of this aphasia have been mentioned earlier (Gorno-Tempini et al., 2004, 2008). Such patients should therefore appear in the present cohort, assuming that LPA accounts for a substantial proportion of progressive aphasic cases. No patient fitted this pattern precisely (Figs 1 and 2). Patient D fitted closely except for an increased frequency of grammatical errors, while Patients O and M fitted slightly less well (poor performance on the WCST, with either speech sound errors or normal phrase length), and Patient L less well again (speech sound errors, with errors on the WCST and Camel and Cactus Test and normal TROG score), demonstrating a gradual shading-off from prototypical LPA into classical PNFA.

There have also been reports of patients developing an isolated progressive impairment of motor speech control, with articulatory distortion and/or groping causing dysprosodic speech, but minimal impairment of language function as such (Broussolle et al., 1996; Duffy, 2006). Insofar as our study assesses these features, Patient I fitted this pattern closely, except that one in five of his utterances was grammatically incorrect. Patients G and H also showed a predominantly speech motor disorder, but they too made grammatical errors and also showed impairments on cognitive testing. Our data therefore suggest that, like LPA, this pattern merges into PNFA as a whole rather than forming a discrete subgroup.

### Table 2 Significant correlations between the variables

<table>
<thead>
<tr>
<th>Correlation (r)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Significant after Bonferroni correction</strong></td>
<td></td>
</tr>
<tr>
<td>Phrase length</td>
<td>Elliptical phrases</td>
</tr>
<tr>
<td>Speech sound errors</td>
<td>Elliptical phrases</td>
</tr>
<tr>
<td>Speech sound errors</td>
<td>Grammatical errors</td>
</tr>
<tr>
<td>Speech rate</td>
<td>Elliptical phrases</td>
</tr>
<tr>
<td>Speech rate</td>
<td>Phrase length</td>
</tr>
<tr>
<td><strong>Significant using uncorrected alpha</strong></td>
<td></td>
</tr>
<tr>
<td>Speech sound errors</td>
<td>Phrase length</td>
</tr>
<tr>
<td>Phrase length</td>
<td>Subordinate phrases</td>
</tr>
<tr>
<td>Elliptical phrases</td>
<td>Grammatical errors</td>
</tr>
<tr>
<td>Speech sound errors</td>
<td>Speech rate</td>
</tr>
<tr>
<td>TROG</td>
<td>WPMT</td>
</tr>
<tr>
<td>Camel and Cactus</td>
<td>WPMT</td>
</tr>
<tr>
<td>Camel and Cactus</td>
<td>WCST</td>
</tr>
<tr>
<td>Camel and Cactus</td>
<td>TROG</td>
</tr>
<tr>
<td>Graded naming</td>
<td>TROG</td>
</tr>
</tbody>
</table>

**Regional atrophy on MRI**

The rating scale was designed such that a score of $≥ 2$ indicates clinically significant atrophy. Reassuringly, 85% of the areas in control scans were rated as 0 or 1, and the others not $> 2$.

As expected, the patients’ scans were more atrophic as a group: 41% of the ratings were at 0 or 1, 48% at 2 and 11% at 3. A rating of 4, which signifies virtual absence of the structure in question, was assigned to only one area for one patient.

Figure 3 summarizes the distribution of atrophy in patients and controls. Areas from adjacent brain regions have been averaged: frontal convexity with insula in ‘frontal’, all temporal neocortical regions (superior, middle and inferior gyri, anterior to posterior) in ‘temporal’, all hippocampal regions (anterior, middle and posterior) in ‘hippocampal’, and all other medial temporal regions (the anterior parts of the parahippocampal gyrus, collateral sulcus and fusiform gyrus) in ‘parahippocampal’. The figure illustrates that the atrophy was not solely frontal, as sometimes has been suggested: both the temporal neocortex and the hippocampus were each affected to a similar extent across the group as a whole. Many of the patients had striking hippocampal atrophy, a finding previously described in other FTLD variants (Davies et al., 2004; Barnes et al., 2006; van de Pol et al., 2006), but not in PNFA. Other medial temporal areas, typically affected in SD (Davies et al., 2004), were less often abnormal in our patients. Additionally, although it was not included in the rating scale, an informal inspection of the scans suggested significant bi-parietal atrophy in at least three patients.

The left and right hemispheres were each more atrophic in patients than in controls, but not necessarily equally so. Although the rating scale is an ordinal variable, the sum of the ratings in one hemisphere relative to the sum in the other hemisphere can be used, with caution, as an indication of asymmetry. Figure 4 shows that more patients than controls were positioned some distance from the dashed line representing symmetry.
Although all the patients were right handed, two had more right-
than left-sided atrophy.

Among the patients, the left and right hemispheric sum scores
are precisely uncorrelated, with \( r = -0.02 \). One could therefore
tentatively suggest that while bilateral atrophy is the rule in
these patients, left- and right-hemisphere deterioration may
proceed at more or less independent rates. The left-sided atrophy
sum score was associated with both Graded Naming and the
TROG \( r = 0.70 \) and 0.65, respectively), but none of the speech
variables were significantly associated with either left- or right-
sided atrophy. The sample size did not give sufficient power to
comment on the associations between individual brain areas and
speech or other cognitive variables.

Discussion

In this study, we have described in detail the spontaneous
conversational speech of a group of patients with progressive
aphasia, broadly defined but excluding those with SD. We have
also explored the relationships of a variety of speech characteristics
with both specific cognitive deficits and regional brain atrophy.
The consensus criteria (Neary et al., 1998) describe non-fluent
speech with anomia, phonemic paraphasia and/or agrammatism.
In the present study, a moderate excess of speech sound and
grammatical errors was observed in almost every case, and
spoken syntax was simplified but rarely frankly telegraphic. Slow
speech and short phrases were also characteristic, although not
universal. Picture naming, syntactic comprehension and executive
function were impaired across the group. The data, therefore,
support the view that a single syndrome, PNFA, accounts for at
least a large majority of progressive aphasic patients without SD.
Patterns resembling LPA and progressive apraxia of speech were
observed, but such patients appear as points at the edges of a
space of continuous variability within a coherent PNFA syndrome.
Despite this consistency in the clinical features, the distribution
of atrophy was remarkably variable—frontal and/or temporal
patterns were observed, symmetrical or asymmetrical. The hippo-
campus was often also affected.

Speech characteristics

Where direct comparisons are possible, our control values align
well with those in previous studies, supporting the validity
of our methods (Table 1). For example, a study using picture description (Graham et al., 2004) reported a control mean of 137 words per minute for speech rate, an open-class word proportion of 44%, a noun/verb ratio of 1.29 and a speech sound error rate of ~1%. This can be compared with a study of narrative speech (Thompson et al., 1997) which reported an open-class word proportion of 48%, a noun/verb ratio of 1.01 and observed that ~25% of control utterances were not full grammatical sentences.

With respect to patients’ performance, the most consistent abnormalities were reduced frequencies of subordinate and relative clauses and increased frequencies of speech sound and grammatical errors. It is, however, important to emphasize that high error rates were rare. All but three of the patients produced phonemically correct words at least 90% of the time, and grammatically acceptable phrases at least half of the time. This finding may help resolve an apparent contradiction between two claims in the literature regarding PNFA: first, that it is characterized by grammatical errors (Turner et al., 1996; Thompson et al., 1997; Neary et al., 1998; Gorno-Tempini et al., 2004); and second, that such errors are not common (Graham et al., 2004). Both claims appear to be true: in this study of 15 progressive aphasia cases, an augmented rate of grammatical errors was an almost universal feature, but generally at the level of occasional inaccuracies rather than a pervasive agrammatism.

It has previously been argued that PNFA differs in certain important respects from Broca’s aphasia (Graham et al., 2004; Patterson et al., 2006). This finding is confirmed here: unlike Broca’s aphasics, these patients did not differ substantially from their matched controls in the number of arguments per verb, the proportion of inflected verb tokens, the frequency of contracted forms, the proportion of open-class words or the noun/verb ratio. It is also noteworthy that misuse or substitution of closed-class words, sometimes said to be characteristic of PNFA (Mesulam et al., 2003), was infrequent, never more than 1 in 35 such words in this study.

There was a substantial but not universal reduction in phrase length, comparable with the widely quoted ‘mean length of utterance’. This phenomenon is probably best seen as a manifestation of simplified grammar, as well as increased failure to complete a phrase. Speech rate is widely reported to be reduced on a group level in PNFA, forming part of the justification for the label, and this too was the case here. Once again, however, the reduction was profound only in two or three cases, and three patients actually had speech rates within normal limits; although the variation amongst the controls on this measure was also striking. Elicitation methods which constrain output, such as picture description, may lead to a more marked difference from control speech rates, as would be expected if word retrieval deficits account for a significant part of the reduction.

Neuropsychological impairments

A reliable impairment across the group was found on three tests: the WCST, Graded Naming and the TROG. Performance on the Camel and Cactus Test was frequently, but only mildly impaired, only occasionally and very mildly on the WPMT; and the VOSP-PD was not significantly impaired. It is of particular interest that patients selected on the basis of a relatively pure language impairment are so often, and so severely, impaired on the WCST; this has been noted before (Broussolle et al., 1996; Nestor et al., 2003). Executive and working memory impairments probably also account for the (mild) impairment observed on the Camel and Cactus Test.

Impairment in the comprehension of complex syntax is well-described in PNFA and confirmed here, though it is interesting that the score on the TROG did not correlate significantly with any of the syntactic production measures in this study. This may reflect a genuine dissociation of syntactic production from comprehension; alternatively, performance on the TROG may be influenced by other factors such as difficulty retaining the spoken sentence during the task—working memory again (Grossman and Moore, 2005). In discussing working memory, it may be relevant to mention its most widely quoted metric, namely the digit span. Although this was not recorded as part of the present study, we have data for 13 of the patients, of whom eight (62%) had a forward span of four or fewer digits, and nine (69%) had a backward span of three or fewer digits.

Relationships between the variables

There were strong correlations between most of the key speech variables; speech rate, phrase length and elliptical phrase frequency were particularly strongly associated. It seems reasonable to assume that the fewer words patients can produce in a given time, the less likely they will be to construct lengthy sentences, and the more often they will resort to isolated phrases to express their meaning. The rates of speech sound and grammatical errors were also highly correlated, but these were less strongly associated with the other speech variables, accounting for some of the observed heterogeneity in the speech of PNFA patients.

Significant associations were also found within the set of cognitive tests, but these impairments did not correlate with the severity of the spoken language impairments. This lack of association has been described previously (Rogers and Alarcon, 1998). Apart from picture naming, the tests in this study did not demand spoken output, and they assess different abilities from those measured by spontaneous speech. On the other hand, it is perhaps surprising that associations are found between tests of cognitive skills as separate as naming and syntactic comprehension, while no such association is seen with the severity of the impairment in the patients’ conversation.

It may be that the particular tests used here were all affected by the same underlying cognitive deficit; as discussed above, working memory is one possible candidate. Such a deficit might also be associated with impaired retrieval of phonological word forms, affecting picture naming. Moreover, a deficit in auditory–verbal short-term memory has been reported in the absence of noticeable speech impairment (Warrington et al., 1971; Saffran and Marin, 1975; Shallice and Butterworth, 1977), reinforcing the observed dissociation between speech and other cognitive impairments.
Symptom profile in individual cases

Some researchers have suggested that progressive aphasic patients outside of SD should be split into distinct subgroups. In particular, LPA and progressive apraxia of speech have been proposed as syndromes categorically distinct from PNFA. When the cases in the present study were compared with the descriptions of these subgroups in the literature, a gradual shading-off from near-prototypical cases through intermediate profiles to typical PNFA was found. This suggests the possibility that LPA and progressive apraxia of speech do not form well-defined subgroups of progressive aphasia, but rather represent points at the edges of a space of continuous variability within PNFA.

Regional atrophy

A number of studies have assessed the distribution of grey matter atrophy in PNFA using volumetric analysis of MRI scans (Nestor et al., 2003; Gorno-Tempini et al., 2004, 2006; McMillan et al., 2004; Josephs et al., 2006). In these group analyses, almost all the affected areas are in the left hemisphere; the inferior frontal gyrus and the anterior insula are consistently abnormal, but atrophy has also been reported in other frontal and temporal areas, as well as in the caudate nucleus.

In the present study, both frontal and temporal areas were highlighted, as well as the hippocampus. The definitions of PNFA used in the earlier studies were generally fairly specific, often requiring motor speech impairment and/or agrammatism, while our inclusion criteria were broader (though still excluded SD). Nonetheless, some of our patients with prototypical PNFA showed clear temporal as well as frontal atrophy.

While left-sided atrophy clearly predominated in the group as a whole (all of whom were right handed), the opposite pattern was seen in two patients. These Patients (G and I) were notable for their relatively pure motor speech impairment. Right-predominant atrophy (in right-handed patients) is well-recognized in the context of this type of presentation (Soliveri et al., 2003; Vitali et al., 2004; Josephs et al., 2006).

Homogeneity and heterogeneity: making sense of PNFA

One of the aims of this study was to determine the similarities and differences between patients with progressive aphasic syndromes other than SD. We deliberately defined the patients more broadly than the consensus criteria for PNFA, in order to assess whether a consistent syndrome of PNFA would emerge.

The results identified a wide variety of impairments on the three levels—conversational speech, cognitive testing and regional brain atrophy—with many of these common to a large majority of the patients, including slow, simplified speech with moderately increased speech sound and grammatical errors, and impairments in executive function, syntactic comprehension and naming, associated with fronto-temporal atrophy. It therefore seems reasonable to treat this patient group as a coherent diagnostic unit.

Nonetheless, considerable variability was observed within the group, in accordance with the variety of clinical descriptions in the literature. Some of this variability resolves under correlational analysis into three partially independent clusters of deficits: (i) slow speech with simplified and elliptical grammar; (ii) speech sound and grammatical errors; and (iii) working memory and executive impairments.

This formulation helps to resolve the apparent contradiction, within the existing literature and also highlighted in the present study, between uniformity and heterogeneity. A wide range of severity was observed on each of the three deficit clusters, with some patients nearly normal and others profoundly impaired. However, this variation is not associated with a single severity factor. Rather, each cluster of deficits has its own degree of severity, which is to some extent independent of the others. In this way, some double dissociations were observed between the deficit clusters; for example, patients with either mild or severe speech disturbance could have either mild or severe deficits on cognitive testing. Similar comments apply to slow and simplified speech on the one hand, and speech errors on the other.

Such a situation, where each patient shows deficits in a range of domains but with variable severity in each, is apt to give the clinical impression of several distinct groups of patients. This is because cases are observed in whom one or other deficit cluster predominates, and these cases are considered prototypical for separate syndromes. Indeed, the profiles of LPA and progressive speech apraxia may represent such cases. The more detailed analysis presented here, however, reveals such patterns as domains of impairment, all of which are typically affected to varying degrees in any given case. The observation that prototypical cases of, for example, LPA are the exception, while overlap cases are the rule, tends to confirm this view. On the other hand, cases traditionally considered prototypical for PNFA tend to be those who do show overlap. In fact, in the consensus criteria for PNFA, the core features cover all three axes: non-fluent speech, agrammatism, phonemic paraphasias and anomia. This is therefore a ‘prototype’ of rather a different sort from LPA or pure speech apraxia.

It has been argued that splitting off typical cases of one form or another has predictive value for post-mortem pathological findings (Josephs et al., 2006; Mesulam et al., 2008). However, recent work has also demonstrated that most patients progress from an initially pure syndrome to further domains of impairment, and that the pattern of disorders, as they appear sequentially through a patient’s illness, has strong pathological predictive value (Kertesz et al., 2005, 2007). One may reasonably predict, therefore, that as more detailed longitudinal phenotypic descriptions of PNFA are brought together with post-mortem data, it will be the patterns and proportions in which the features co-occur, which will associate with pathological subtype.

The results of this study have demonstrated a coherent syndrome, which is common to patients presenting with progressive aphasia, excluding SD. The term ‘progressive non-fluent aphasia’, as defined by the widely accepted consensus criteria (Neary et al., 1998), accurately describes most (if not quite all) of these patients, and a reduction in speech fluency is
certainly a typical and salient feature of the group; this term therefore remains the most appropriate designation for the syndrome.

Acknowledgements

We are grateful to Rhys Davies and Chris Kipps for help and advice with the MRI rating scale, to Tomasz Bak for the TROG data, to two anonymous reviewers for their perceptive and helpful comments, and most importantly to the patients and their carers for their unstinting generosity and patience.

Funding

Programme Grant from the Medical Research Council (62825, to J.R.H.); Interdisciplinary Behavioural Sciences Center Grant from the National Institute of Mental Health (MH64445, to K.P.).

References


