Lesion correlates of patholinguistic profiles in chronic aphasia: comparisons of syndrome-, modality- and symptom-level assessment

Ilona Henseler,1,2 Frank Regenbrecht2 and Hellmuth Obrig1,2

One way to investigate the neuronal underpinnings of language competence is to correlate patholinguistic profiles of aphasic patients to corresponding lesion sites. Constituting the beginnings of aphasiology and neurolinguistics over a century ago, this approach has been revived and refined in the past decade by statistical approaches mapping continuous variables (providing metrics that are not simply categorical) on voxel-wise lesion information (voxel-based lesion–symptom mapping). Here we investigate whether and how voxel-based lesion–symptom mapping allows us to delineate specific lesion patterns for differentially fine-grained clinical classifications. The latter encompass ‘classical’ syndrome-based approaches (e.g. Broca’s aphasia), more symptom-oriented descriptions (e.g. agrammatism) and further refinement to linguistic sub-functions (e.g. lexico-semantic deficits for inanimate versus animate items). From a large database of patients treated for aphasia of different aetiologies (n= 1167) a carefully selected group of 102 first ever ischaemic stroke patients with chronic aphasia (>12 months) were included in a VLSM analysis. Specifically, we investigated how performance in the Aachen Aphasia Test—the standard clinical test battery for chronic aphasia in German—relates to distinct brain lesions. The Aachen Aphasia Test evaluates aphasia on different levels: a non-parametric discriminant procedure yields probabilities for the allocation to one of the four ‘standard’ syndromes (Broca, Wernicke, global and amnestic aphasia), whereas standardized subtests target linguistic modalities (e.g. repetition), or even more specific symptoms (e.g. phoneme repetition). Because some subtests of the Aachen Aphasia Test (e.g. for the linguistic level of lexico-semantics) rely on rather coarse and heterogeneous test items we complemented the analysis with a number of more detailed clinically used tests in selected mostly mildly affected subgroups of patients. Our results indicate that: (i) Aachen Aphasia Test-based syndrome allocation allows for an unexpectedly concise differentiation between ‘Broca’s’ and ‘Wernicke’s’ aphasia corresponding to non-overlapping anterior and posterior lesion sites; whereas (ii) analyses for modalities and specific symptoms yielded more circumscribed but partially overlapping lesion foci, often cutting across the above syndrome territories; and (iii) especially for lexico-semantic capacities more specialized clinical test-batteries are required to delineate precise lesion patterns at this linguistic level. In sum this is the first report on a successful lesion-delineation of syndrome-based aphasia classification highlighting the relevance of vascular distribution for the syndrome level while confirming and extending a number of more linguistically motivated differentiations, based on clinically used tests. We consider such a comprehensive view reaching from the syndrome to a fine-grained symptom-oriented assessment mandatory to converge neurolinguistic, patholinguistic and clinical-therapeutic knowledge on language-competence and impairment.

Keywords: lesion-deficit analysis; VLSM; chronic aphasia; aphasia syndromes; linguistic modalities
Lesion correlates of patholinguistic profiles in chronic aphasia

Introduction

Grounded in Broca’s seminal discovery of a relationship between lesion topography and language impairment, further differentiation between aphasia syndromes has been prompted by the description of quite remarkable dissociations between impairments affecting different linguistic levels and modalities in aphasia (Ardila, 2010). These clinically observed dissociations informed models, which conceptualize language competence as the orderly interplay between separable modules affording specific linguistic tasks (Heilman, 2006). However, despite progressive sophistication of such ‘box-and-arrow’ models the approach has been challenged (Poeppel and Hickok, 2004); not all clinically observed symptomology can be accommodated and correlations between lesions in ‘classical’ language areas and respective syndromes turned out to be brittle (Mohr et al., 1978; Dronkers et al., 2007). Moreover, functional imaging of linguistic competence in the intact brain delineated a widespread bilateral large scale network by far outweighing classical language areas (Sidtis, 2006), while at the same time a given cortical area was shown to be involved in quite different linguistic processes (Vigneau et al., 2006).

In view of such potential shortcomings of previous model-based approaches a voxel-by-voxel analysis has been proposed (Bates et al., 2003), yielding a statistical measure of how strongly a lesion in a given voxel predicts impairment in a specific language task. Voxel-based lesion–symptom mapping (VLSM) avoids methodological limitations of some previous approaches for lesion-deficit analysis. Instead of an a priori selection of patients based either on lesion pattern or on patholinguistic profile it does not proceed from a topographic or patholinguistic hypothesis, but uses voxel-wise lesion information and the full spectrum of behavioural data. This allows for a detailed analysis of the relationships between brain damage and performance. Applying statistical principles developed for functional brain imaging, VLSM has the advantage to identify regions that are critical for a certain function, rather than all regions potentially recruited by a task and identifies these in grey and white matter. Finally, it allows for regressing out the influence of factors, such as language-related cognitive functions or lesion size, to provide a fine-grained picture of the functional significance of particular brain areas.

In the past decade this potent tool has been successfully applied in a number of studies on patients with language impairment. Studies on verbal fluency (Bates et al., 2003; Baldo et al., 2006), sentence comprehension of varying syntactic complexity (Dronkers et al., 2004; Magnusdottir et al., 2012), speech production (Dronkers and Ogar, 2004; Borovsky et al., 2007), noun versus verb naming (Piras and Marangolo, 2007), verbal and non-verbal semantics (Aires et al., 2009; Schwartz et al., 2009; Walker et al., 2011), picture naming (Rudrauf et al., 2008; Cloutman et al., 2009; Parkinson et al., 2009; Piras and Marangolo, 2010), and word and number reading (Piras and Marangolo, 2009), and speech repetition (Fridriksson et al., 2010) provide an increasingly detailed map of areas mandatory for specific linguistic sub-functions. Additionally the overlap between linguistic and more general cognitive processing has been targeted (Leff et al., 2009) and subcortical connections constituting the network have been investigated (Turken and Dronkers, 2011; Kummerer et al., 2013). Interestingly, however, the relation between lesion pattern and the ‘standard’ aphasic syndromes has as yet not been formally addressed by VLSM. This may reflect a general skepticism with regard to the validity of clinical syndrome classification, a notion shared by language therapists who call for symptom- rather than syndrome-guided intervention. An early CT-based attempt supports such skepticism since retrospective analysis of 221 patients did not show a relationship between lesion pattern and syndromes (Willmes and Poeck, 1993). Additionally, automated and probabilistic syndrome classification is not provided by most test batteries preventing a graded measure of syndrome classification. Such graded measures are the prerequisite to exploit the pivotal advantage of the VLSM over other lesion-symptom approaches. Hence an integrative view correlating lesions to different levels of aphasia classification is lacking.

The present study aims to extend previous research in targeting exactly this. Specifically, we used the data from the standard clinical test battery for aphasia in German (the Aachen Aphasia Test, AAT) to examine the relationship between brain damage and a broad spectrum of language functions in a large sample of patients with chronic aphasia as a result of first ever ischaemic stroke, including small, circumscribed but also large territorial lesions. Since this tool for chronic aphasia assessment not only includes tests covering different language modalities at all linguistic levels but also allows for probabilistic syndrome classification, impairment can be directly compared at the syndrome-, modality-, and even more fine-grained specific symptom-level in the same group of patients. Although the terms ‘modality’ (e.g. repetition, comprehension, etc.) and linguistic level (e.g. ‘lexico-semantics’) refer to largely accepted concepts in patholinguistics, a general term for the more fine grained test and functional level is lacking. To refer to this level for the AAT subtest (e.g. repetition of sentences) and the additional tests (e.g. synonym judgement) we therefore use the term ‘specific symptoms’ throughout the text. Besides the different levels of assessment provided by the AAT, we included results from additional, very specific language and neuropsychological tests—completed by subgroups of patients—to find out whether these would help to refine and complement the results of the AAT-based lesion–symptom analysis.

The analyses reported in this paper comprise three parts. First, VLSM maps for the four standard syndromes (Broca’s, Wernicke’s, global and amnestic aphasia) were generated. Second, we identified brain lesions significantly associated with impairments in the linguistic modalities assessed with the AAT, including the different assessment levels in spontaneous speech and the paper-and-pencil tests. Third, additional language tests targeting more specific

Abbreviations: AAT = Aachen Aphasia Test; RWT = Regensburger Wortflüssigkeits Test; VLSM = voxel-based lesion–symptom mapping; WMS-R = Wechsler Memory Scale-Revised

by guest on 08 May 2018
symptoms in mildly affected patients were analysed and neuropsychological scores were factored out to further refine the lesion–symptom analysis. Thus the present study provides comprehensive data on the associations of scores from standard language testing, clinical syndromes, and specific language tests with the location of brain lesions in patients with aphasia. We hypothesized that such a comprehensive view at differentially fine grained levels of classification may be more adequate to illustrate the complex interplay between linguistic levels and modalities, parametric increase in recruitment with increasing complexity of the test material and syndromes relating to a vascular territory, at the basis of most classification attempts.

Materials and methods

Subjects

Patient data were retrieved from the database of the Max-Planck-Institute for Human Cognitive and Brain Sciences and the Clinic for Cognitive Neurology in Leipzig, Germany. The clinic treats patients with acquired brain lesions due to a wide range of aetiologies in a day-care setting. The setting plus the research-orientation and a clinical focus on professional re-integration skew the cohort towards lesser affected and younger patients. The initial search yielded 1167 patients who had undergone extensive language testing; 1093 were excluded, because they did not fulfil our inclusion criteria, which were: (i) single left-hemispheric ischaemic stroke in the middle cerebral artery territory; (ii) right handedness; (iii) native speakers of German; (iv) normal language development; (v) no concomitant severe microangiopathy; (vi) no dementia or other neurodegenerative disease; (vii) no pre-stroke history of psychiatric or neurological disease; and (viii) high resolution MRI available. One hundred and two patients fulfilled these criteria. Mean age of these participants was 52 years [standard deviation (SD) = 12 years] and mean time post-stroke was 12.6 months (SD = 20.0 months; see Supplementary Table 1 for more details on the patient cohort). Data were collected in accordance with the Declaration of Helsinki and the ethics committee of the University of Leipzig. Based on the syndromatic classification implemented in the AAT, the sample included the following aphasia types: Broca (n = 16), Wernicke (n = 9); global (n = 8); amnestic aphasia (n = 28). Patients in whom the ALLOC classification yielded a probability of <70% were regarded as non-classifiable (n = 7) and patients in whom classification probability for aphasia was <80% were regarded as residual (n = 28). Few patients showed language impairment according to clinical judgement but scored within normal limits in the AAT (n = 6). The AAT classifies conduction and transcortical aphasias according to operational criteria beyond the ALLOC classification. This applied to only four patients in our sample and is hence not investigated further. Overall severity of the impairment according to the Token Test included severe (n = 8), moderate (n = 11), mild (n = 31), and minimal (n = 49) impairment. Thus patients showed a wide range of aphasic deficits and varied with regard to severity as required for the VLSM analysis. A lesion overlay map of the patients is displayed in Fig. 1. Comparisons with the standardization cohort (n = 376) of the AAT (Huber et al., 1983) showed similar mean age and duration of the aphasia, whereas our cohort included a larger portion of less affected patients by comparison to the standardization cohort (Supplementary Table 1).

Behavioural tests

Patients were examined with a large battery of language and neuropsychological tests by professional speech therapists and neuropsychologists (Table 1 and Supplementary material). The primary language measure was the AAT (Huber et al., 1984), the standard diagnostic battery for chronic aphasia in Germany. The AAT is a multi-dimensional diagnostic tool allowing for syndrome classification (Albert et al., 1981) and reliable assessment of all clinically relevant language modalities. Specifically it comprises a rating of spontaneous speech along six dimensions in a semi-standardized interview, four fully standardized tests and the Token Test. In detail spontaneous speech is scored for (i) communicative abilities; (ii) articulation and prosody; (iii) phonematic structure; (iv) semantic structure; (v) syntactic structure; and (vi) speech automatisms. The standardized tests include: (i) speech repetition (phonemes, 1-syllabic words, compound words, foreign words, sentences); (ii) written language (writing, combing words, reading); (iii) picture naming (simple objects, colours, compound words, situations); and (iv) language comprehension (auditory words, auditory sentences, reading words, reading sentences). The Token Test (Orgass et al., 1973) is included as a complex measure of overall auditory language comprehension and is typically used to classify overall aphasia severity. Apart from the comprehensive coverage of patholinguistic levels the AAT has a special feature in that it allocates patients to one of four major syndromes (Broca, Wernicke, global and amnestic) with a probability in per cent. This syndromatic classifier based on a non-parametric discriminant analysis procedure (Habbema et al., 1974) renders the likelihood of the respective syndrome (e.g. a patient may be classified 90% Broca, 10% amnestic). Patients in whom the probability of any of the four syndromes is

Figure 1 Lesion overlay. Lesion overlay map of the 102 patients, showing the coverage of most of the left hemisphere. Coloured areas are lesioned in at least one patient. Voxels lesioned in >10% of all patients are depicted in yellow. Left hemisphere is left on coronal slice. Numbers refer to MNI coordinates.
Lesion correlates of patholinguistic profiles in chronic aphasia

Brain 2014: 137; 918–930

Those identified as 'unclassifiable' are provided. The four main syndromes according to the ALLOC procedure of the Aachen Aphasia Test are indicated in bold. The values given in the following 6 columns represent the percentage of patients (% of respective tests) who could be allocated to one of the four main syndromes according to the ALLOC procedure of the Aachen Aphasia Test. Additionally, percentages of patients classified as residual or no aphasia and those identified as 'unclassifiable' are provided.

### Table 1 Behavioural measures

<table>
<thead>
<tr>
<th>Test</th>
<th>Patients (n)</th>
<th>Broca</th>
<th>Wernicke</th>
<th>Amnesic</th>
<th>Global aphasia</th>
<th>Residual / no aphasia %</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT</td>
<td>102</td>
<td>16</td>
<td>9</td>
<td>27</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>Spontaneous speech scoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Articulation and prosody</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phonemic structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syntactic structure</td>
<td>102</td>
<td>16</td>
<td>9</td>
<td>27</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automated speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formalized subtests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Speech repetition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Written speech</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Picture naming</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Language comprehension</td>
<td>84</td>
<td>19</td>
<td>11</td>
<td>32</td>
<td>7</td>
<td>24</td>
</tr>
<tr>
<td>Token test</td>
<td>102</td>
<td>16</td>
<td>9</td>
<td>27</td>
<td>8</td>
<td>34</td>
</tr>
<tr>
<td>Additional tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific semantic tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semdiff nouns</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>74</td>
<td>7</td>
</tr>
<tr>
<td>Semdiff verbs</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>64</td>
<td>9</td>
</tr>
<tr>
<td>Lemo25 (synonym judgement)</td>
<td>10</td>
<td>30</td>
<td>0</td>
<td>60</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Specific picture naming tests</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-time critical</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-animate</td>
<td>30</td>
<td>13</td>
<td>7</td>
<td>37</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>-inanimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-verbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-compounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time-critical, RAPID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-animate</td>
<td>26</td>
<td>4</td>
<td>0</td>
<td>19</td>
<td>69</td>
<td>8</td>
</tr>
<tr>
<td>-inanimate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-verbs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-compounds</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word fluency tests (RWT)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lexical word fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic word fluency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic category shift</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>52</td>
<td>12</td>
</tr>
<tr>
<td>Neuropsychology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WMS-R (span forwards/ backwards)</td>
<td>78</td>
<td>13</td>
<td>4</td>
<td>32</td>
<td>3</td>
<td>41</td>
</tr>
</tbody>
</table>

Table listing all tests used for the VLSM analyses (for more details please see Supplementary material). Absolute numbers (n) of patients who completed the respective tests are indicated in bold. The values given in the following 6 columns represent the percentage of patients (% of n) who could be allocated to one of the four main syndromes according to the ALLOC procedure of the Aachen Aphasia Test. Additionally, percentages of patients classified as residual or no aphasia and those identified as ‘unclassifiable’ are provided.

<70% are usually regarded unclassifiable (but see below for inclusion of these patients in the VLSM approach).

In contradistinction to the syndrome-oriented approach of the AAT, we included a number of language tests focussing on more specific aspects of language dysfunction (for details see Table 1 and Supplementary material). These tests were completed in smaller subgroups of the sample studied. The choice of the different tests was guided by clinical judgement and served patholinguistic differentiation mostly in patients with mild impairment. There are tests targeting the semantic differentiation of nouns and verbs (‘Semdiff nouns’, ‘Semdiff verbs’, in-house tests), a synonym judgement test (LeMo subtest 25), tests for picture naming (a large naming battery, including a number of different categories, tested in a normal and a time-critical condition), and tests for word fluency (Regensburger Wortflussigkeits Test, RWT).

From the large neuropsychological test battery we included a measure of the verbal working memory capacity (Wechsler memory scale revised, WMS-R) to factor out the impact of this language-related cognitive function on the results of the analyses.

### Imaging methods and voxel-based lesion–symptom mapping analysis

For all patients, structural high-resolution MRI scans were available from a 3 T Siemens MRI system (Siemens Trio or Verio system, Siemens Medical Systems) including 3D T1-weighted- (1 mm isovoxel), and FLAIR images. For VLSM analyses, stroke lesions were manually drawn in all three planes (axial, coronal, sagittal) on each slice of the T1 images using MRICron (Rorden and Brett, 2000), as it has been shown to be the best method for lesion mapping as compared with semi-automated or automated processing (Wilke et al., 2011). FLAIR images were used as a reference. This was done by a trained neurologist (I.H.) with extensive experience in clinical neuroradiology and checked by a second experienced neurologist (H.O.). Both raters were blind to the individual’s symptoms, when reaching consensus on lesion delineation. Next, using SPM8 (www.fil.ion.ucl.ac.uk/spm), images were transformed into standard stereotactic space (MNI, Brett et al., 2001) and segmented by the unified segmentation approach (Ashburner and Friston, 2005), which is robust, even in the case of large lesions (Crinion et al., 2007). Estimation of normalization parameters was restricted to healthy tissue applying the predefined lesion mask (i.e. cost function masking; Brett et al., 2001). The resulting normalized lesion maps were then transferred into ‘NPM’ [non-parametric mapping, implemented in MRICron; (Rorden et al., 2007), www.mricron.com]. The VLSM approach applies the graded information at the behavioural level to voxel-wise binary lesion information. For each test, the patients’ lesion maps and the corresponding behavioural scores are used. In a voxel-by-voxel approach patients are divided into two groups according to whether or not the respective voxel is lesioned. Performance scores are statistically compared between groups by T-statistics as has been recently recommended by Medina et al. (2010). To grant reasonable levels of statistical power, statistical analyses were restricted to those voxels where at least 10% of the patients were lesioned. The resulting maps were corrected for multiple comparisons using a 1% false discovery rate (FDR corrected \( P < 0.01 \)) for the AAT subtests. Since the additional specific language tests were only administered to subgroups of our sample (for details see Table 1), we used a more liberal FDR of \( P < 0.05 \) in the respective analyses. Only voxels exceeding these thresholds are shown in the maps below.

### Results

The results of our VLSM analyses are reported starting with the syndrome level to then present results on linguistic levels/modalities targeted by subtests of the AAT. Finally we present the findings.
regarding specific symptoms. Analyses at these three levels were performed in the majority of patients based on the AAT assessment. For the specific symptom level, results of the additional tests assessing language in subgroups of patients (Table 1) are reported in a separate paragraph. At the end of this section we briefly report on correlations between scores of the different tests and subtests.

**Syndrome level: Aachen Aphasia Test and ALLOC classification**

VLSM analyses for Broca’s and Wernicke’s syndromes yielded anterior and posterior lesion sites, respectively. Notably the lesion maps for either syndrome showed minimal overlap (Fig. 2). In detail, lesions correlating with the score for Broca’s aphasia were located in left inferior frontal gyrus (including Broca’s area), the left insular cortex, as well as the adjacent white matter and basal ganglia. In contrast, lesions associated with Wernicke’s aphasia were located in the left posterior and anterior temporal and parietal cortex (including Wernicke’s area). Note that this analysis used the graded probabilities from the syndrome classifier in that a patient with probability scores of for instance 50% Broca, 30% global and 20% amnestic aphasia would be regarded ‘unclassifiable’ in AAT assessment, but would relevantly contribute to the VLSM maps for all three syndromes. The probability scores for global aphasia correlated with lesions partially overlapping with those for Broca’s and Wernicke’s aphasia but extending to the left supramarginal gyrus and angular gyrus, as well as to the left superior temporal cortex and the left inferior and middle frontal cortex (Supplementary Fig. 2). The latter were located more anterior than lesions associated with Broca’s aphasia. VLSM analysis using the classifier for ‘amnestic’ aphasia showed no consistent statistically significant lesion cluster (data not shown).

**Dimensions of spontaneous speech and the Token Test**

VLSM analyses revealed coherent and statistically significant lesion-performance associations for some of the scores regarding linguistic levels of spontaneous speech. In particular, performance on the level ‘articulation and prosody’ in spontaneous speech significantly correlated with damage to the left insula and the adjacent opercular cortex, and the putamen (Fig. 3). In other words, damage to these structures yielded significantly higher scores for articulatory impairment. For the level ‘phonematic structure’, there was a significant association between behavioural impairment and damage to the left supramarginal gyrus (Fig. 4), i.e. patients with damage to this area presented significantly more phonematic errors in spontaneous speech than patients with intact
supramarginal gyrus. In contrast to these circumscribed performance-lesion relations, scorings of more complex characteristics of spontaneous speech were related to less specific lesion patterns. Impairment on the level ‘semantic structure’ was associated with lesions in a large area encompassing the temporoparietal, anterior temporal, and inferior frontal cortex, whereas scoring for ‘syntactic structure’ in spontaneous speech were mainly affected by lesions to the posterior and middle part of the superior temporal gyrus (Fig. 5). The scorings for ‘communication’ and ‘speech automatisms’ correlated to unspecific perisylvic lesions, suggesting that these scores either represent quite general impairment in aphasia or that the distribution of the scores in our sample was strongly skewed due to the inclusion of rather mildly affected patients. Similarly correlation with performance in the Token Test yielded a large uncharacteristic lesion area.

**Standardized Paper and Pencil Tests**

Significant associations between impaired performance and distinct lesion sites were also found for subtests of the AAT, which allow for numerical rating of the linguistic modalities: (i) speech repetition; (ii) naming; (iii) written language; and (iv) language comprehension. For ‘speech repetition’ damage to the left supramarginal gyrus and the left posterior superior temporal gyrus correlated with lesser performance. Refining the analysis based on the scores for the sub-tests (targeting specific symptoms in the respective modality at different linguistic levels) revealed that gradually more areas come into play during the repetition of phonemes, mono-syllabic words, compound nouns, and sentences. This suggests that damage to the left supramarginal gyrus is associated with impaired repetition on all levels, while repetition of more complex material (e.g. sentences) requires integrity of additional regions in its vicinity extending into the posterior superior temporal gyrus (Fig. 5). For the AAT-subtest ‘language comprehension’, VLSM analysis revealed that damage to more anterior brain regions are crucial for this modality as assessed by the AAT. Critical areas included the anterior and middle part of the left superior temporal gyrus and in the left anterior inferior frontal gyrus (Fig. 6). In contrast with these circumscribed lesion patterns for speech repetition and language comprehension, reduced performance in the AAT-subtests ‘written language’ and ‘picture naming’ was related to relatively unspecific areas of damage.

**Tests beyond the Aachen Aphasia Test: lexico-semantics and fluency**

More specialized tests were performed in smaller subgroups of patients in whom the linguistic deficit required further differentiation based on clinical judgement (Table 1).

As opposed to the rather unspecific results when using the AAT picture-naming subtest, a large picture naming battery (170 items) suggests that integrity of the left orbital inferior frontal gyrus (Brodmann area 47) and the posterior middle temporal gyrus are most critical for naming (Fig. 7). Interestingly the comparison of different semantic categories yielded different lesion foci: naming of non-living compared to living objects more strongly depended on the integrity of the middle temporal cortex (Fig. 7). Conversely, both categories showed significant overlap in the frontal cortex. Another interesting distinction emerged in that naming of compound when compared with single nouns correlated with relative integrity of the left anterior temporal lobe. Not regarding primarily linguistic aspects of naming, an additional analysis demonstrated that dissociable areas are pivotal for picture naming in a ‘normal’ (naming without time restraint) and a time-critical condition. In contrast to normal naming, speeded naming significantly depended on the integrity of the left posterior inferior frontal gyrus and the left inferior frontal junction area (Fig. 7B).

Figure 8 summarizes the results for additional tests evaluating lexico-semantic capabilities. Lesions in the orbital part of the inferior frontal gyrus correlated with performance in a synonym judgment test (Fig. 8A). Clinical tests for semantic differentiation developed in-house allowed for a finer-grained mapping; firstly larger parts of the inferior frontal gyrus seem to be necessary for semantic differentiation of verbs versus nouns. Additionally, lesions in the orbital inferior frontal gyrus showed higher correlation with the semantic differentiation of nouns, whereas a more posterior, triangular/opercular part of the inferior frontal gyrus was associated with the impaired semantic differentiation of verbs (Fig. 8B versus C). For executive requirements inherent to fluency tasks as tested by the German standard tool assessing word fluency (RWT), fronto-opercular and insular areas adjacent to the left inferior frontal gyrus come into play. Comparison between critical areas for semantic fluency and categorical shift...
requirements (Fig. 8D versus E) further illustrate the increasingly distributed network responsible for increasingly complex lexico-semantic tasks. (cf. Fig. 7 suggesting partially overlapping relevance of the inferior frontal gyrus for both lexico-semantic tests.)

Influence of verbal working memory

One critical factor for most auditory language tests is verbal working memory (Leff et al., 2009). Therefore we regressed out verbal working memory performance in the VLSM analysis (this was done in a subgroup of patients; Table 1). As an example, the analysis of the AAT subtest for speech repetition with verbal working memory as covariate revealed that the integrity of the left posterior superior temporal gyrus is the most critical region for unimpaired speech repetition (Fig. 5B). The AAT analysis of the subtest written language (subtests writing by dictation and combining words by dictation) showed that, after factoring out verbal working memory, the region around the left angular gyrus remained the most critical region for this modality.

Behavioural correlations for subtests within the AAT and additional tests

We calculated correlations between all subtests of the AAT and correlated the results of the additional tests with selected lexico-semantic scores as assessed by the AAT (see Supplementary Tables 2 and 3). The AAT subtests generally showed high
intercorrelations (Pearson’s $r > 0.5$, often $> 0.7$) with the exception of ‘articulation and prosody’ scored for spontaneous speech and the formalized subtest ‘auditory sentence comprehension’. As expected, correlations of the syndrome probabilities and the subtests were low and inconsistent. In sum the data are in line with the correlational analysis reported in the original construction sample of the AAT (Huber et al., 1983) confirming that the ALLOC classification discriminates syndromes based on a highly correlated subtest data space.

Regarding the additional tests performed in a smaller subset of patients (Supplementary Table 3) the most notable finding was that scores for ‘regular’ unconstraint naming did not correlate with those for naming of the same items under time constraint. This held even within most categories (e.g. ‘plants’). Interestingly time-constraint naming correlated with digit span backward (WMS-R bw) and semantic fluency (RWT), suggesting that time-constraint naming may tax more executive rather than primarily linguistic functions.

### Discussion

Syndromic classification of aphasias is rooted in the tradition to make inferences about linguistic modules by observation of patholinguistic dissociations. However, the quest for precise localization of such linguistic modules in the brain has only been feasible since high resolution neuroimaging provides functional and structural information intra vitam. Here the rapidly growing body of functional imaging studies (Price, 2010) discloses that the network affording language competence by far exceeds the ‘classical’ perisylvian language areas. Conversely most of the ‘key language areas’ have been shown to be involved in quite divergent, partially non-linguistic tasks (Vigneau et al., 2006). Therefore correlating high-resolution lesion imaging with performance has become an invaluable approach to complement functional imaging results. It may help to differentiate between mandatory brain areas and those facultatively involved in dissociable aspects of language processing. Proceeding from the many studies published in the past decade, which correlate specific partial experimental tests with lesion pattern, we here addressed the question of how far clinically applied diagnostic dimensions can be ‘mapped’ to circumscribed brain structures. Importantly we included different levels of patholinguistic characterization asking whether and how syndrome-, modality-, and more specific symptom- level analyses will converge or diverge regarding the lesion patterns seen in chronic aphasia. Although such an approach is exploratory by default, our expectation was that the coarse and
controversial syndrome-based classification would yield rather un-
specific lesion patterns, while a finer grained picture was expected
when mapping test results targeting specific modalities and symp-
toms. Such expectation rests on early work attempting syndrome-
lesion correlation (Willmes and Poeck, 1993) and on the largely
accepted notion that the complexity of linguistic operations cannot
be modelled in a limited small number of underspecified modules
forming the basis for syndromatic aphasia classification (Poeppel
and Hickok, 2004).

Lesion correlates of syndrome
classification

The most remarkable finding of our analyses is that, contrary to
our expectation, VLSM of the AAT syndrome classification yielded
a clear-cut distinction between lesion sites for ‘Broca’ and
‘Wernicke’ aphasia (Figs 2 and 9). ‘Broca aphasia’ correlated
with lesions in the left inferior frontal cortex, the left insular
cortex plus adjacent white matter and basal ganglia. Wernicke’s
aphasia was associated with lesions in left temporal (including
superior middle and inferior portions, superior temporal gyrus/
middle temporal gyrus/inferior temporal gyrus), temporo-parietal
(temporo-parietal junction/inferior parietal cortex) and temporo-
ocipital cortex. There was no significant overlap between these
VLSM patterns. The result is notable form two perspectives. First,
from a clinical-neurological vantage point it suggests that
parametric classification according to the ALLOC procedure imple-
mented in the AAT analysis quite precisely dissociates an anterior
from a posterior aphasia syndrome. As the AAT was constructed
and validated on cohorts of patients with stroke (Supplementary
Table 1) it highlights that aphasia syndromes have been largely
shaped by the most common cause of aphasia, that is, ischaemic
stroke. Hence they must be considered partially aetiology-specific,
pertaining to the territories of mostly pre-rolandic versus superior
temporal artery. The distinction between two aphasic syndromes,
roughly corresponding to a Wernicke- and Broca-type aphasia
runs like a common thread through the aphasia literature (Tesak
and Code, 2008); however, this first statistical imaging approach
regarding syndrome classification highlights that vascular anatomy
is an important aspect constituting the distinction. Second,
the results may be equally informative from a neuro-linguistic perspec-
tive. They indicate that despite a relative specialization of brain
areas for dissociable linguistic processes, lesions in a highly and
reciprocally connected network are likely to interfere with a
number of linguistic functions. Additionally co-occurrence of
top-down and bottom-up processes is uncontroversial to support
uncompromised linguistic competence, thus a spectrum of linguistic
(i.e. a ‘syndrome’) rather than a selective breakdown of a ‘linguis-
tic module’ can be expected even for highly localized lesions.

It should be noted that the clear distinction of an anterior from
a posterior aphasia syndrome does not imply that all ‘Broca’ and
‘Wernicke’ apahas can be expected to show lesions in the re-
spective territory. The AAT yields pseudo-continuous probabilities
for syndrome classification based on the multidimensional subtests.
Thus in the individual patient patholinguistic features of different
syndromes may overlap, indicating partial or overlapping lesions,
due to stroke aetiology or individual vascular anatomy. Therefore
the delineations of the syndromes in Figs 2 and 9 must be con-
sidered the statistically derived maximal extension of a common
underlying territory. Conversely it must be noted that the results
do not allow for a delineation of the minimal lesion extent exten-
datory to produce a Broca or Wernicke aphasia according to the
diagnostic criteria of the AAT. Although the study did include
patients with smaller (i.e. more strategic) lesions, the statistical
approach constituting VLSM requires lesion overlap in a suffi-
ciently large number of patients (statistical power will be highest
when 50% of the patients in a respective voxel are lesioned). Such
methodological constraints and the fact that we investigated
chronic aphasia may favour larger lesions to contribute to the
overall results and may sharpen the distinction between the
large Broca- versus Wernicke-related lesion areas. To answer
the question whether a given area is mandatory to cause a specific
syndrome, work on (sub)acute stroke has been shown to be
highly informative (Hillis et al., 2006). Indeed a study comparing
acute and chronic stroke elegantly showed that the association
between lesions in Broca’s area and Broca’s aphasia was only sig-
nificant in acute perfusion-based lesion analysis (Ochfeld et al.,
2010).

In contrast with the clear border between an anterior and pos-
terior aphasia syndrome, global aphasia correlated with lesions
extending into the anterior frontal, superior temporal and tem-
poro-parietal cortex (Supplementary Fig. 2). Unsurprisingly there
was substantial overlap with the above two syndromes. Although
our sample showed a bias for mild to moderate aphasia, we sug-
gest that the finding indicates that severe impairment in global
aphasia will not allow for more specific lesion-behaviour correl-
ations. Finally amnestic aphasia, the fourth syndrome whose like-
ilhood is predicted by the AAT classifier, showed no coherent
lesion pattern, in line with the fact that amnestic aphasia is
diagnosed in patients with a great variety of different residual symptoms.

**Differentiation according to linguistic modalities**

Syndrome classification in the AAT is afforded by a non-parametric discriminatory procedure (Habibena et al., 1974) converging scaled semi-standardized ratings of six dimensions of spontaneous speech, four fully standardized subtests and the Token Test (Orrgass et al., 1973). Going to this next finer scale, VLSM yielded distinct areas in which a lesion predicts more severe impairment of the respective modality. Notably distinct lesion patterns were seen already regarding some dimensions (based on linguistic levels) which are scored for the modality ‘spontaneous speech’. Interestingly this relevance of the semi-standardized spontaneous speech analysis corresponds to the predictive power of these ratings with regard to the syndrome classifier. As noted in the initial construction and validation of the AAT (Huber et al., 1984), the acuity of syndrome classification is only slightly enhanced by the four formalized subtests, whereas the Token Test is largely used for the description overall severity of the aphasia.

Concerning the evaluation of spontaneous speech, the most circumscribed VLSM results were obtained for the dimension ‘articulation’ and ‘phonematic structure’. Articulatory deficits corresponded to lesions in the left insular cortex and the adjacent opercular and subcortical structures (Figs 3 and 9). This finding is in agreement with previous findings strongly suggesting that the insular cortex is the most important brain region supporting complex articulatory movements (Dronkers, 1996) and has been highlighted to be the most relevant lesion site to correlate with apraxia of speech (Dronkers and Ogar, 2004; Ogar et al., 2006). Additionally, insular cortex has been shown to be critical for the articulatory tasks in patients with aphasia, especially, in words of high articulatory complexity (e.g. more syllables or initial consonant cluster; Baldo et al., 2011). The current results fit well in this picture. As no formal testing but semi-standardized rating of spontaneous speech is used for this AAT-score it suggests that across tasks and notably also in spontaneous speech, articulatory performance strongly relies on the left insula.

Another dimension of spontaneous speech rating, ‘phonematic structure’, yielded circumscribed VLSM results in the left supramarginal gyrus (Fig. 4). Phonological processing involves several parts of the language network (Vigneau et al., 2006). Interestingly our lesion data converge with a ‘virtual lesioning’ of the supramarginal gyrus with transcranial magnetic stimulation. The latter study showed that inhibition of the supramarginal gyrus interferes with performance in a phonological task (Hartwigsen et al., 2010). As will be discussed below, our results further indicate that damage to the left supramarginal gyrus affects the ability to repeat speech.

Speech repetition has been shown to largely rely on the dorsal pathway (Hickok and Poeppel, 2004; Saur et al., 2008). The present VLSM analysis of the respective AAT subtest suggest left supramarginal gyrus, left posterior superior temporal gyrus and the adjacent white matter to be critical for this linguistic modality (Fig. 5). This finding in chronic aphasia nicely complements reports on the association between impaired repetition with cortical hypoperfusion in the inferior portion of the left supramarginal gyrus in acute stroke patients (Fridriksson et al., 2010). Most traditional models of aphasia regard the arcuate fascicle essential to repetition tasks and indeed Fig. 5 supplies evidence for its involvement in the sample studied here. Factoring out verbal working memory (Fig. 5B), the arcuate fascicle shows less involvement; however, to disentangle cortical contributions from those caused by lesions to the fibre bundles, fibre tracking approaches will be mandatory as has been shown in acute aphasia (Kummerer et al., 2003). Moving to a finer grained level, greater complexity of the material to be repeated increased the area crucial for repetition performance (phonemes < 1-syllabic words < sentences, Fig. 5A). Yet, it must be taken into account that performance in tests of speech repetition in patients with left posterior superior temporal gyrus/supramarginal gyrus lesions may also be influenced by their impaired verbal working memory capacity. Indeed recent studies have demonstrated that there is an overlap of both functions on the neuroanatomical level (Leff et al., 2009; Buchsbaum et al., 2011). Therefore we regressed out verbal working memory performance in an additional VLSM-analysis (Fig. 5B). This analysis supplied a more circumscribed lesion pattern, suggesting that the integrity of the left posterior superior temporal gyrus, the left inferior supramarginal gyrus, and the adjacent white matter is most critical for the repetition aspect.

In contrast with the dorsal pathway of language, which is thought to support audio-motor mapping, the ventral pathway may be crucial for the extraction of meaning during speech comprehension (Hickok and Poeppel 2004; Saur et al., 2008). It comprises more anterior temporal regions, including anterior and middle portions of the superior temporal gyrus, the inferior frontal cortex, and the capsula extrema. Largely consistent with this concept of a ventral stream, the current VLSM-maps (Fig. 6) revealed that lower scores for ‘language comprehension’ are associated with damage to the left superior temporal gyrus/middle temporal gyrus, extending to anterior temporal areas and into the inferior frontal gyrus.

The score for ‘syntactic structure’ of spontaneous speech is the most explicit rating of syntactic abilities in the AAT. Though this is quite an unspecific measure for complexity of propositional speech, VLSM analysis regarding this dimension of spontaneous speech in our sample converged on a posterior temporal and parietal area (Supplementary Fig. 1). This is remarkable as the inferior frontal gyrus (‘Broca’s area’), posterior superior temporal gyrus and in part, the anterior temporal lobe have all been shown to contribute to syntactic processing although their respective roles remain quite controversial (Humphries et al., 2006; Grodzinsky and Santi, 2008; Tyler et al., 2011). The rather coarse rating of syntactic abilities in the AAT certainly does not allow us to disentangle differential contributions but may point to the concept that posterior superior temporal gyrus affords integration of syntactic and semantic information (Friederici, 2012). Such successful integration may be a crucial factor determining structural complexity of utterances in spontaneous speech.
Lesion correlates of additional specific tests

The AAT-based assessment of ‘semantic structure’ in spontaneous speech and the formalized subtest ‘picture naming’ only yielded relatively broad perisylvic lesion patterns. Therefore we additionally performed VLSM with clinically used tests probing lexico-semantics on a finer scale. It should be noted that these tests were administered to smaller subgroups of the patients (Table 1). Therefore a comparison between the AAT-based assessment at one of the three levels and the results pertaining to the additional tests needs to acknowledge that clinical judgement introduces a bias. Patients examined by these tests typically had a mild or residual aphasia (see Table 1 for percentages of syndromes in the respective subgroups). Respecting this caveat the additional tests on semantic differentiation of nouns and verbs and synonym judgement identify the left inferior frontal gyrus as a core region for the processing of lexico-semantical information. The patients’ ability to differentiate between semantically related words was found to rely heavily on the integrity of this area, supporting the notion that left inferior frontal gyrus is critical for strategic processing of semantics.

Additionally, inferior frontal gyrus could be shown to play an important role in executive aspects of language at the lexico-semantical level, as tested by semantic word fluency and semantic category shift (Fig. 8) in line with previous reports assessing verbal fluency (Baldo et al., 2006) but also conversational speech (Borovsky et al., 2007).

Using a large battery assessing confrontational naming (Fig. 7) VLSM delineated a network including left middle temporal gyrus, the left anterior temporal lobe, and the left anterior inferior frontal gyrus. The left middle temporal gyrus and anterior inferior frontal gyrus have previously been observed in functional MRI studies to be sensitive to semantic priming (Gold et al., 2006) and to semantic relations (Thompson-Schill et al., 1997; Wagner et al., 2001; Snyder et al., 2007). Regarding lexico-semantics, there is converging evidence for the left superior and middle temporal gyrus and the anterior temporal lobe to house the lexicon (middle temporal gyrus; Baldo et al., 2013) and to take part in supramodal semantic integration (anterior temporal lobe; Patterson et al., 2007). Correspondingly, lesions in these temporal areas are associated with more semantic errors in naming tests (Piras and Marangolo, 2007, 2010; Schwartz et al., 2009; Walker et al., 2011) and with a reduced type/token ratio in spontaneous speech, reflecting reduced semantic variability (Borovsky et al., 2007). As proposed by early work (Vandenbergh et al., 1996) a recent study elegantly used lesions-based information to delineate fibre pathways connecting the left anterior inferior frontal gyrus with the left middle temporal gyrus, underlining that a semantic processing relies on a highly connective fronto-temporal network (Turken and Dronkers, 2011).

Completed by a smaller subgroup of mild to moderately impaired patients, specific clinical tests allowed us to differentiate between semantic categories (Fig. 7A): anterior temporal lesions more severely impaired naming of living items, whereas impaired naming of non-living objects correlated with posterior temporal damage, in line with previous functional imaging data (Damasio et al., 1996). Another notable result is that naming under time critical conditions requires the integrity of a different set of brain areas. The respective VLSM maps (Fig. 7B) suggest that posterior inferior and middle frontal brain regions, including the inferior frontal junction, are especially important for picture naming under time pressure. These brain areas have been shown to be essential for executive and attention-related tasks (Milham et al., 2003; Brass and von Cramon, 2004). The result highlights the importance of cognitive control in a task quite challenging even for mildly impaired patients with aphasia and are of note regarding the clinical relevance of such tests. We suggest that time-constraints in a language test may shift the focus from a purely linguistic to a more verbal-executive domain. This is also supported by the behavioural data in our sample, which showed low correlations between regular and time-constraint naming, while the latter correlated with more executive-linguistic functions such as verbal working memory and semantic fluency (Supplementary Table 3 for details).

Although some of the specific tests were performed in rather small subsamples of the 102 patients and should hence be regarded with caution, the results presented here are in line with traditional lesion studies. As most of these rather demanding tests cannot be performed in severe aphasia, future work should address how such fine-grained differentiation of mostly residual language impairment relates to a more fine-grained map of lesion distribution.

Limitations and perspectives

Lesion based maps of linguistic functions (Fig. 9) underestimate the fact that language competence relies on highly connected sub-functions both linguistically and functional-anatomically (Friederici, 2012). Additionally patients with chronic aphasia modify communicative behaviour over time and the network affording linguistic processing will show partial functional adaptation to the lesion as has been shown for the (sub)acute stage (Saur et al., 2006). Instead of a clear-cut correlation between a certain cortical area and a linguistic function, lesion-symptom correlations must hence be considered a starting point to investigate structural and functional changes in aphasia. Whether or not such plasticity partially resembles the life-long changes in the unimpaired language network is another extremely exciting challenge for future research.

Our results show that distinct brain lesion patterns can be delineated for different levels of clinically applied patholinguistic classification in chronic aphasia. The non-overlapping dissociation between an anterior/posterior aphasia syndrome corresponding to the classifiers differentiation between ‘Broca’s’ and ‘Wernicke’s’ aphasia may be somewhat unexpected but strongly underlines the relevance of vascular territories regarding syndrome classification. Future studies will need to address the question in how far aetiology and the respective lesion topography will influence VLSM results. In this vein a recent study demonstrated topographical and patholinguistic differences in semantic dementia and chronic ‘Wernicke’s’ aphasia (Ogar et al., 2011). At the level of the AAT-based rating of spontaneous speech and the modality-specific subtests showed circumscribed lesion patterns which...
Lesion correlates of patholinguistic profiles in chronic aphasia Brain 2014: 137; 918–930 | 929

Partially overlapped (Fig. 9). Such overlap and the fact that some subtests contributing to the syndrome classification did not yield circumscribed lesion patterns highlights the importance to consider networks rather than specific areas to afford linguistic competence for specific linguistic levels and modalities. Here tractography (Turken and Dronkers, 2011; Kummerer et al., 2013) and further functional activation studies in the lesioned network will be required to address compensatory and adaptive changes in the respective networks especially when chronic aphasia is under investigation (Price and Crinion, 2005). For lexico-semantic capacities the necessity of large and categorically organized test batteries is illustrated by the comparison of the AAT test ‘picture naming’ and more specific fine-tuned test batteries. Though clinical practice may require some parsimony for realistically applicable testing, specific symptoms may require a much more fine-tuned instrumentation to allow differentiation of deficits as the basis for deficit-adapted therapy. We regard such sharpening of diagnostic tools in aphasia diagnosis equally important for the much under-determined linguistic level of syntax. This is a challenging but exciting task with the potential to disentangle functional-structural correlates of syntactic impairment, a differentiation not afforded by most clinical test batteries.

Supplementary material

Supplementary material is available at Brain online.

References


Ardila A. A proposed reinterpretation and reclassification of aphasic syndromes. Aphasiology 2010; 24: 363–94.


