Published in final edited form as:
*Neuroimage*. 2013 June ; 73: 200–207. doi:10.1016/j.neuroimage.2012.09.007.

# Neuroimaging in aphasia treatment research: Issues of experimental design for relating cognitive to neural changes

**Brenda Rapp**[1], **David Caplan**[2], **Susan Edwards**[3], **Evy Visch-Brink**[4], and **Cynthia K. Thompson**[5]

[1]Johns Hopkins University, Department of Cognitive Science, Baltimore, Maryland, USA [2]Harvard Medical School, Department of Neurology, Boston, Massachusetts, USA [3]University of Reading, Department of Psychology, Reading, UK [4]Erasmus University, Medical Centre, Department of Neurology, Rotterdam, The Netherlands [5]Northwestern University, Department of Communication Sciences and Disorders, and Neurology, Evanston, Illinois, USA

## Abstract

The design of functional neuroimaging studies investigating the neural changes that support treatment-based recovery of targeted language functions in acquired aphasia faces a number of challenges. In this paper, we discuss these challenges and focus on experimental tasks and experimental designs that can be used to address the challenges, facilitate the interpretation of results and promote integration of findings across studies.

### Keywords

Aphasia; Language recovery; fMRI

## 1.0 Introduction

Functional neuroimaging techniques provide exciting new opportunities to further our understanding of the neural substrates and neural changes that support treatment and recovery in cases of acquired language disorders, but the design of these neuroimaging investigations presents a number of challenges. This paper addresses several issues of experimental design, focusing on issues of task selection and development for purposes of evaluating language and cognitive functions, on the one hand, and evaluation of neural activation and changes, on the other. It is not a review of the literature or a cookbook for designing specific experiments, but rather a discussion of some key experimental design issues that arise in research of this type and suggestions for addressing them.

There are different theoretical frameworks for aphasia treatment and the framework adopted will have specific implications for the design and interpretation of neuroimaging studies of aphasia treatment. One major distinction among aphasia treatment approaches is between those that target specific aspects of language difficulties and those that target



Correspond with: Brenda Rapp: rapp@cogsci.jhu.edu.

communication skills more broadly (Byng & Duchan, 2005). The former type of approach assumes that improvement of deficient language operations will lead to improvement in communication and quality of life, while the latter assumes that the communication system and, therefore, the overall language network should be the target of treatment (see Martin, Thompson, & Worrall, 2008, for in depth discussion of these approaches and their integration). To date, most studies of the neural changes that support aphasia treatment and recovery have been undertaken within the "language functions" framework, and in this paper we shall limit our discussion to studies of this sort

There have been several excellent papers dealing with experimental design issues concerning neuroimaging of aphasia treatment (e.g., see Meinzer, Harnish, Condway & Crosson, 2011; Thompson & den Ouden, 2008; Crosson et al., 2007; Crinion & Leff, 2007; Price & Crinion, 2005). In this paper we integrate some of the key insights from these previous papers and we focus on those aspects of experimental design that we consider to be important for facilitating the interpretation of the complex data sets generated by research on this topic.

## 2.0 The Goals of Neuroimaging of Aphasia Treatment Research

Neuroimaging studies of aphasia treatment have largely addressed three types of research goals. One type of goal, and the one that will form the primary focus of this report, is to characterize and understand the neural changes that support treatment-induced improvement of specific language functions[1]. It is widely accepted that different language functions recruit different neural regions (e.g., thematic role assignment processes in sentence comprehension are likely supported by different neural substrates from those involved in the retrieval of phonological word forms from a lexical phonological memory store) and thus it is likely that there are differences in the neural topography of recovery for different language functions. The information gained from this type of research will contribute importantly to our understanding of the functional organization of the brain and of its plastic and dynamic properties vis a vis language.

A second type of goal of neuroimaging in aphasia treatment research is to characterize neural changes that result from treatments that directly target specific neural substrates. Examples include pharmacological treatments such as those targeting dopamine receptors (Bachman & Morgan,1988), or techniques such as repetitive transcranial magnetic stimulation (rTMS) that may be used to target specific brain regions in order to facilitate recovery or therapeutic outcomes (e.g., Naeser, Martin, Nicholas, Baker at al., 2005). Functional imaging has rarely been used to evaluate the neural changes that occur in these types of treatment (e.g., Martin, Naeser, et al., 2009), but many of the issues concerning the characterization of functional (psychological) changes that occur as a result of a treatment and the associated neural changes would be the same as those that we discuss in the body of this paper.

Finally, a third type of research goal may be to identify neural factors that predict therapeutic outcomes. These neural factors may be useful in predicting the effectiveness of treatment for specific language disorders or may have more general predictive value. This type of information would be of considerable clinical significance with implications for prognosis, treatment design and even allocation of resources. Some research with a predictive goal may be highly similar to that carried out for the previously described goal of understanding brain changes that underlie recovery of language functions. That is, it may well be that an understanding of the neural changes that support recovery of specific

[1]By language or cognitive "functions" we mean what are sometimes referred to as processes or operations.

language functions would contribute to understanding the conditions that may be necessary to achieve therapeutic success for specific language deficits. However, research with a predictive goal may also involve functional neuroimaging of quite a different nature. We discuss issues related to this third type of research goal briefly here.

The goal of identifying neural factors that are predictive of therapeutic outcomes affects the design of neuroimaging studies differently than does the goal of identifying neural changes associated with recovery of specific language functions. This is because, essentially, any type of neural measure could turn out to be important for predicting the effectiveness of treatment: measures of brain structure, resting metabolic, hemodynamic or physiological states, and other measures that do not involve activation of the brain by specific stimuli or tasks, might well turn out to be predictive of treatment effects. For instance, the extent of structural damage or hypometabolism in the dominant hemisphere, the most common frequency in a resting awake EEG or characteristics of the brain's default network (see Buckner, Andrews-Hanna, & Schacter, 2008) might be extremely valuable for predicting the expected course (natural history) of aphasia as well as an individual's response to treatment. That is, individuals with larger lesions, less metabolic activity, and slower EEG rhythms may be more neurologically impaired, have less undamaged brain tissue available for recovery, and may respond less well to treatment. Along these lines, Thompson, Den Ouden, Bonakdarpour, et al. (2010) found higher perfusion levels (i.e., greater blood flow values) in regions of the brain showing post-treatment upregulation of neural activation, in a study of six agrammatic aphasic patients. The hemodynamic response function time-to-peak (HRF TTP) also was faster in these regions. Or, in a recent study by Meinzer et al. (2010), it was suggested that in aphasic individuals with a stroke in the middle cerebral artery, the degree of integrity of the hippocampus and the surrounding white matter might be crucial for the success of anomia training. In addition, it may be the case that the neural responses to very general language tasks (e.g., of the sort used in fMRI studies designed to replicate Wada testing or as language localizers; Bookheimer et al, 1997; Fedorenko et al, 2009) serve to identify brain regions with potential to support language recovery.

Despite the wide range of measurements that could be made in research studies with a predictive goal, or perhaps because of this wide range, these studies face their own specific challenges. First, expertise is important in determining which observations are most likely to be useful. Second, it is clear that most neural observations will be imperfect predictors of outcome, and that it will be necessary to consider multiple factors (e.g., age, handedness, gender, education, premorbid language proficiency (especially in polyglots)) to gain predictive power. Clearly, a research program directed at the question of which neural measures predict recovery and response to treatment is highly complex. The point here is simply that many of the issues we raise in the body of the paper may not be key considerations for research with the goal of identifying neural factors predictive of treatment success.

We now return the discussion specifically to issues regarding the goal of identifying and understanding the neural changes and conditions that support recovery of specific language functions and response to treatment. The experimental design of these studies must be such that the study can support both the interpretation of the findings within the context of the study, and also the integration of the findings of the study with other relevant studies within the broader literature. The organizing principle for this type of research is that it involves relating changes in language/cognitive function to changes in neural substrates such that the language functions and the neural substrates are, essentially, the two sides of an "equation" that map one on to the other. This type of research most typically asks: If a treatment is effective in changing certain aspects of an individual's language function, what neural changes support that improvement? In order to identify and quantify changes in language

functions and neural substrates, experimental tasks must be selected and/or developed that will effectively shed light on both the cognitive and neural substrates. Thus, task development and experimental design are critical to the success of this type of research.

The burgeoning field of cognitive neuroscience is directed largely at understanding the relationship between cognitive functions and neural substrates in the intact brain. However, in research on the neural substrates of aphasia treatment the complexity of the research is considerably increased by the addition of: (a) disruption of normal language operations and the neural substrates that support them, and (b) changes from pre- to post-treatment at both cognitive and neural levels.

## 3.0 Characterizing the Disrupted Language System

Research examining the effects of treatment on recovery of particular language operations requires the identification, prior to treatment, of the language functions that have and have not been disrupted by the lesion. Similarly, subsequent to treatment, it is necessary to identify both the language functions that have improved and those that have not. The need to characterize the nature of the language deficit/s applies not only to studies of the neural correlates of treatment but also to studies directed at identifying neural predictors of recovery.

Evaluations of language functions are guided by some (explicit or implicit) theory of language. For example, in research involving acquired reading deficits it would be important to know if the investigators assume multiple routes for reading aloud and, if so, what operations are assumed to be carried out by these routes. Or, in the case of deficits in spoken word naming, theoretical assumptions regarding stages of spoken word production are critical for the selection and interpretation of diagnostic tasks.

Characterization of the language profile will involve the characterization of deficient and intact language and related cognitive functions based on the theoretical framework of the investigator. This can be carried out with different degrees of specificity. For example, in the domain of reading, a reading deficit may be described broadly as an acquired dyslexia, or more precisely depending on the investigator’s assumptions about routes for reading aloud and the operations that are assumed to be carried out by these routes. For example, a particular individual’s deficit may be described, fairly precisely, as a disruption of the processes involved in converting sublexical orthographic units to phonological units.

The degree of specificity that is sought will determine the tasks that are administered. In the case of reading, a sublexical orthography-to-phonological conversion deficit requires documentation of a characteristic profile on a number of tasks, such as difficulty in nonword reading, reduction in regularity effects in reading and lexical decision, normal frequency effects in word naming and lexical decision, etc. Changes associated with treatment must be documented at the same level of detail. Interpretation of findings can never be more specific than the specificity with which the language profile has been characterized. For example, neural changes observed subsequent to treatment for dyslexia cannot be attributed to changes in sublexical conversion if the diagnostic work was not carried out at this level of specificity. Furthermore, the manner in which results can be integrated across studies is affected by the compatibility of the different studies in terms of the specificity with which the language profiles are characterized. That is, it is difficult, if not impossible, to integrate studies that broadly test for and treat dyslexia with those that identify and target specific subcomponents of the reading process. In general, questions regarding general aspects of language functions require general language measures and questions regarding specific language functions require more detailed measurements. The matching of the granularity

(level of detail) between research questions, language and neural measurements and data interpretation applies broadly throughout this type of research.

There also is a need to achieve comprehensiveness in the diagnostic evaluation. This is important in order to understand the full impact of the treatment; that is, the degree to which treatment impacts other cognitive systems associated with language. If the diagnostic tasks only characterize the language impairment, it will not be possible to determine if treatment affects only language functions or if it has a broader impact. Understanding both the specific and more general consequences of treatment is also critical for interpreting the neural changes. For example, if changes from pre- to post-treatment are noted in both nonword reading and verbal working memory, then the neural changes observed cannot be interpreted as simply or solely reflecting changes in the neural substrates supporting sublexical reading processes. Additionally, this type of broad range of testing may also be useful for research directed at identifying language/cognitive profiles predictive of therapeutic success.

To date, a great deal of research on the neural substrates of aphasia treatment has studied individuals with spoken naming deficits (see Thompson & Den Ouden, 2008, for review). Thus, we illustrate the points discussed above by briefly reviewing a study reported by Menke, Meinzer, Kugel, Deppe et al. (2009). These researchers were interested in examining the neural changes that support successful treatment for anomia. They examined eight individuals with chronic spoken language production impairments and carried out fMRI scanning before and immediately after a 2-week intensive spoken naming treatment as well as at an 8-month follow-up. An extensive pre-treatment language and cognitive evaluation was administered. Results showed that all participants suffered from a deficit affecting the linking of semantic information with its corresponding phonological word form. Although not explicitly stated, we can assume that diagnostic testing was based on the theoretical view that the spoken word production system consists of multiple stages of processing that allow word semantics, lexical retrieval, phonemic encoding and motor production to be distinguished from one another (e.g., for a review see Rapp & Goldrick, 2006). Given these assumptions, a disruption in retrieval of word forms from semantic representations will manifest itself with a characteristic profile of impaired and spared performance. On this basis, the diagnostic work carried out by Menke et al (2009) involved testing hearing, auditory language comprehension, object naming, word fluency, and repetition. Spoken naming was further evaluated to determine error profiles (i.e., if errors were phonemic or semantic, if subjects were responsive to phonological cueing, and whether or not there were signs of apraxia of speech). The reasoning was as follows: intact auditory comprehension rules out a semantic deficit, good repetition and only mild (if any) apraxia of speech rule out motor problems, whereas difficulties with picture naming, decreased fluency, production of semantic errors, and sensitivity to phonemic cueing point to a disruption in the link between word meaning and word form (see Table 1). In addition, a more comprehensive evaluation of related language and cognitive abilities was also carried out. This included tests of written language, short term memory (verbal and visual), general learning capacity (visual and verbal), general intellectual functioning, attention and executive functions.

The diagnostic testing in the Menke et al (2009) study represents a clear example of both considerable specificity in the diagnosis and comprehensiveness in the overall testing. Despite the importance of pre- and post-treatment evaluation of language/cognitive profiles, however, a review of the literature reveals that this is rarely done (but see Rapp & Vindiola, 2005; Kiran & Thompson, 2003). Typically, the comprehensive evaluation is carried out only prior to the intervention to establish a diagnostic profile. The therapeutic program is then applied and success associated with treatment is assumed to reflect an improvement specifically in the disrupted language/cognitive operations identified in the pre-therapeutic

evaluation. In fact, in the Menke et al. study, only a pre-treatment comprehensive evaluation was administered, guiding the researchers to target semantic-lexical links with their treatment. The absence of a comprehensive post-treatment evaluation limits the ability to firmly establish the specific impact of the treatment on the language system. This type of comprehensive pre- and post-treatment evaluation is uncommon in treatment studies where practical considerations may limit the feasibility of extensive testing. Nonetheless, it is important to understand the interpretive limitations this may impose.

## 4.0 Evaluating Neural Changes Associated with Specific Language Processes

Having characterized the language and associated deficits, the next challenge is to identify tasks that activate the neural systems that support the language functions of interest in the intact brain. These systems consist of interacting brain areas that are dedicated to particular language operations. Increasingly, researchers have been interested in identifying both the specialized areas and their effective functional connectivity in particular tasks. Whether the focus of a particular study is on particular specialized areas or on their interactions, it is necessary to have tasks that test the language functions of interest in order to observe changes in neural activation patterns that take place as the targeted language function improves from pre- to post-treatment. This may seem straightforward but, in fact, it is typically extremely complex to develop an experimental design that will recruit and isolate the relevant substrates. In fact, few studies examining the neural correlates of aphasia treatment have used neuroimaging designs that pinpoint the neural changes that specifically reflect treatment of targeted functions. It is much more common for researchers to use tasks that capture a broad range of functions rather than the narrow set that is deficient and/or the set that has been the target of treatment.

One of the main reasons that identifying the specific substrates that support relevant language functions is so challenging is that tasks typically recruit multiple functions, not all of which are of interest. As a consequence, it is not sufficient that the tasks used during the functional neuroimaging recruit the substrates of interest, there must be a method for isolating specific regions from among others that may not be relevant. Before discussing various approaches to this issue, we must point out that whatever the technique that is used, it is important that the techniques can be shown to isolate the substrates of interest reliably and consistently in neurologically intact individuals. For this purpose, there must either be an existing literature demonstrating the reliability of the design in activating the relevant substrates in neurologically intact individuals or control participants should be included in the study. A combination of both is ideal. Including control participants who perform the same neuroimaging tasks as the aphasic participants increases interpretability of the results by allowing the researcher to address key questions such as: Do pre-to-post treatment changes recruit “normally recruited” substrates? If so, is the recruitment to the same, greater or lesser degree than in the intact brain? Is there recruitment of novel regions, not observed in the control participants, such as homologous areas in the opposite hemisphere?

One often-used strategy for isolating neural substrates supporting specific language functions is to compare activations produced by experimental and control (or baseline) tasks. Using this logic of pure insertion involves selecting an experimental task, which includes the functions of interest plus functions that are not of interest, and a control task, which includes the functions that are not of interest (Friston et al., 1996). Although this approach is not uncontroversial, depending on how carefully matched the tasks are, the contrast between experimental and control tasks can potentially isolate the neural substrates supporting the functions of interest (see Crosson et al., 2007; Caplan, 2009, for review of different types of control conditions). For example, Peck, Moore, Crosson, et al. (2004) examined the neural

correlates of an aphasia treatment that was directed at improving production of syntax using an experimental task which required participants to produce passive sentences to describe pictures and two control/baseline tasks: passive viewing of nonsense objects and picture naming. The contrast of the experimental task with the control tasks allowed for the broad identification of neural activity associated with sentence production compared to that involved in visual processing and lexical retrieval. However, even using these control tasks, it is possible that the remaining neural activity could reflect a variety of functions in addition to syntactic encoding (the function of interest), for example, selection of thematic roles. Therefore, in the Peck et al. study, the control conditions did not completely isolate the syntactic processing skill targeted in treatment. As a result, interpretation of changes in activation patterns within the "isolated" substrates from pre- to post-treatment was complicated because it is not clear which of the observed neural changes corresponded specifically to improvement in syntactic processing.

A caveat pertaining to what we have said above is that, although the use of scanner tasks that specifically identify the treatment-targeted skill may be the goal for neuroimaging research in aphasia, the extent to which this is possible to achieve is limited by our current understanding of the brain and language or cognitive processing. That is, for many, if not most, cognitive functions there is no consensus on the tasks/experimental manipulations that best serve to isolate their neural substrates in functional neuroimaging. Given this, many researchers find it useful to have an additional relatively non-specific baseline (such as the nonsense objects in the Peck et al. study) that allows a broad range of areas and functions to enter into the analysis rather than mistakenly "subtracting them out".

In addition to the insertion logic, another method for isolating substrates of interest is to use a parametric design (see Caplan, 2009, for discussion). This approach generally involves contrasting different levels of a parameter using the same task. For example, to identify the neural substrates associated with recovery of syntactic processing in patients with agrammatic aphasia, Thompson, Den Ouden, Bonakdarpour, et al. (2010) used an auditory verification task, randomly presenting sentences with different levels of syntactic complexity (e.g., object clefts, subject clefts, and simple actives) for participants to judge as matching or not matching a simultaneously presented visual scene (e.g., a bride carrying a groom or groom carrying a bride). The rationale was that contrasting activation to items of different levels of difficulty or complexity would reveal neural substrates sensitive to syntactic complexity. In another set of studies, Thompson and colleagues (Thompson, Bonakdarpour, Fix, 2010) examined verb processing using verbs differing in the number of arguments they require. Specifically, one-argument (e.g., laugh), two-argument (e.g., fix), and three-argument verbs (e.g., throw) were presented together with pseudowords in a lexical decision task. The neural areas sensitive to this parametric variation in argument structure complexity (e.g., the greater the complexity, the greater the activation) are candidate regions for the representation and processing of verb argument structure.

We note, however, that the use of a parametric design does not, by itself, guarantee that the language operations that are targeted in treatment will be isolated. This is because other functions may be affected by the same levels of the parameters manipulated. For example, in the case of syntactic complexity, it is possible that sentences that differ in syntactic complexity also differ with regard to their working memory demands. In that case, sensitivity to the levels of the syntactic parameter manipulated may appear in brain regions associated with working memory, in addition to (or instead of) areas involved in processing the syntactic aspects of the sentences (but see Santi & Grodzinsky, 2007). This presents a challenge in identifying which of the observed neural effects reflect which cognitive processes. How this is accomplished will vary depending on the targeted function, sometimes with the necessary information being available in the literature, and other times

requiring a series of studies, each varying parameters that affect different components or stages of processing.

A form of parametric variation that deserves special consideration is the manipulation of the degree of exposure to a stimulus of a particular type. The best example of this is simple repetition, which is associated with suppression of neurovascular responses (presumably due to habituation). This technique of "neural adaptation" has been used productively to identify areas associated with particular types of representations and processes, such as recognition of visually presented words (Dehaene et al, 2001). Variants of simple repetition, such as (semantic) priming, can be used to isolate specific aspects of processing. As with other designs, control of correlated properties of stimuli is needed (e.g., orthographic familiarity must be controlled when lexical status is varied to study lexical activation), but these challenge are no greater for designs of this sort than for others.

## 5.0 Evaluating Neural Activity Associated with Intact Language/Cognitive Processes

Just as in the behavioral evaluation it is important to characterize not only impaired but also spared language/cognitive processes, measurement of neural activity associated with spared processes may also provide valuable information. It is generally assumed that language/cognitive skills that are unimpaired and not targeted by treatment should show neither behavioural improvement nor neural change from pre- to post-treatment, but this assumption may be incorrect: the brain may re-organize after a lesion in ways that change the neural systems that support functions that are unaffected at the behavioural level. Stable activations from pre to post treatment associated with these spared functions would indicate that this is not the case and would provide one part of the evidence that neural reorganization that is observed is, in fact, due to treatment. The tasks used to evaluate spared language/cognitive processes may, at the same time, serve as functional localizers, identifying the location of well-established functions so that the results of interest can be localized not only in terms neuroanatomical landmarks or standardized Talairach (Talairach & Tournoux, 1988) or MNI coordinates, but also in terms of their location relative to functional landmarks. This is especially important in lesioned brains, which often are topographically distorted. For example, although not a training study, Tsapkini, Vindiola and Rapp (2011) compared the activation patterns of an individual with acquired dyslexia and dysgraphia (due to left mid-fusiform resection) with those of neurologically intact individuals. In addition to experimental reading tasks used to recruit and isolate the neural substrates of orthographic processing, tasks were also administered that allowed for the localization of neural areas involved in the processing of houses and faces. This was done because, behaviourally, face and house processing had been shown to be intact in the individual that was the subject of the investigation and, furthermore, because the visual processing of these types of stimuli is known to yield highly reproducible activation patterns in neurologically intact individuals. Specifically, the activation loci for object and face processing are thought to be topographically adjacent to the substrates of the fusiform gyrus that are believed to be amongst those normally involved in orthographic processing (McCandliss et al., 2003). This broader neural evaluation, extending beyond the disrupted orthographic functions, allowed Tsapkini et al. to show that, in this individual, the intact cognitive functions of face and object processing were supported by a normal pattern of brain activation, whereas the disrupted orthographic processes showed clear signs of reorganization.

## 6.0 Scanner Task Considerations

While the previous sections in this paper have dealt with broad issues in experimental design, there are additional issues that arise concerning aspects of the neuroimaging tasks

themselves. One of the main points to consider in designing fMRI tasks is the relative difficulty of the task for the participant (Poldrack, 2000). A second issue concerns the relationship between the task(s) used during treatment and the scanner task(s) selected to evaluate the effects of treatment.

There has been some debate in the literature regarding the issue of the difficulty of the functional neuroimaging tasks/items (Crosson, McGregor, Gopinath, et al., 2007; Thompson, 2005; Thompson & Den Ouden, 2008). The question is: Should the tasks/items be “easy” (within the capacity of the aphasic individual) or “difficult” (beyond his/her capacity)? Several early studies of language recovery in aphasia were conducted with “easy” tasks or items that participants could perform with high accuracy (Cao, Vikingstad, George, Johnson, & Welch, 1999; Mummery, Ashburner, Scott, & Wise, 1999; Ohyama et al., 1996). However, this approach may be problematic because such tasks/items may not (sufficiently) engage the defective language/cognitive process in the scanner, potentially failing to reveal changes that may be induced by the treatment. In contrast, the concern regarding the use of difficult tasks/items that yield poor performance is that these tasks may engage processes or strategies other than those under investigation and also decrease the reliability of neuronal activation (Price & Friston, 2002). In addition, difficult tasks/items may induce increased effort, inattention to the task, as well as emotional responses, such as anxiety and frustration. Thus, the reason that task difficulty is an important issue is that it contributes to the complexity of interpretation of the neural activation patterns.

Various solutions have been proposed to address this issue. One solution is to include both easy and difficult items in lists individually tailored to each participant’s performance (e.g., Menke et al., 2009; Fridriksson, et al., 2006). The use of stimuli that are within the capacity of a neurologically injured individual is similar to the use of materials in studies with neurologically intact individuals who are typically performing at a high level of accuracy. In this scenario, with proper experimental controls, improvement in performance (such as the speed with which a task is performed) can be related to changes in neural responses. Likewise, neural changes associated with the ability to process stimuli that were previously beyond a participant’s capacity can shed light on the neural basis for the recovered/successfully treated stimuli. We note, however, that it is an open question whether the neural changes seen in both situations (easy or difficult stimuli) are the same.

A similar approach, but one which does not require developing stimuli for individual study participants, involves separately analyzing items that yield correct and incorrect responses during the functional neuroimaging task. For instance, several studies examining naming treatment (Fridriksson, et al., 2007; Menke, et al., 2009; Vitali, et al., 2007) have compared activation for pictures named accurately during baseline with those named poorly prior to treatment, but relearned by the end of treatment. Likewise, Fridriksson et al., (2006) used naming accuracy during pre and post treatment scans as a covariate in higher level analyses. Another approach may involve comparing the before vs. after treatment changes for correct versus incorrect items (e.g., Meinzer et al., 2006, Postman-Caucheteux et al., 2010). Indeed, using behavioral measures such as accuracy and reaction time of scanner task performance has been used productively to identify neural changes associated with the treatment-induced language improvement.

Another approach involves the use of implicit tasks that may reflect changing skill levels induced by treatment while requiring a relatively constant level of effort in the pre- and post-treatment periods. Passive tasks can be used for this purpose. For example, when the treatment target is word reading, the task could consist of passive viewing of written words. Neurologically intact individuals will automatically read words, even in passive viewing conditions. For individuals with acquired dyslexia who receive treatment, if the ability to

read words increases from before to after treatment, there will be an increase in automatic reading in this passive task, yielding measurable changes in neural activity while minimizing the pre- to post-treatment changes in the effort, anxiety, or frustration that may be associated with an overt task. A limitation is that passive tasks do not provide behavioral evidence (e.g., accuracy and/or reaction time) regarding what was processed, which can introduce additional interpretive challenges. An alternative to passive tasks are paradigms that measure priming, neural adaptation, or habituation with a relatively easy active task. This type of implicit measurement may reveal behavioral and neural changes from pre- to post-treatment, while also providing an explicit measure of processing accuracy. These types of tasks may have the advantage of avoiding problems associated with difficult tasks/items as well as the interpretive limitations associated with simple passive tasks

Another question that sometimes arises is whether or not the task used in treatment and the task used in scanning should be the same. The majority of functional neuroimaging studies of aphasia treatment have involved treatments for word naming difficulties and, in many of these studies, both the treatment and neuroimaging tasks have involved picture naming. The intuition underlying the use of the same task in treatment and scanning is that, in this way, the scanner task will best capture any treatment effects. However, there is no reason why the same tasks must be used. This is especially evident given that the target of the remediation is typically a language function and not a task. For example, the target of remediation may be lexical retrieval while the task used to measure change may be picture naming. In that case one can imagine that a task such as semantic cueing may be used in remediation while picture naming (and relevant control tasks) may be used during scanning. Simply put, the appropriate tasks will be those that, with regard to treatment, best serve to remediate and, with regard to scanning, best serve to recruit and isolate relevant neural substrates. Thus the relationship between tasks employed in treatment and scanning will depend entirely on the particular questions under investigation. There may be multiple processes associated with naming that the experimenter may wish to investigate prior to and following treatment. For instance, Sebastian and Kiran (2011) used two scanner tasks –picture naming and semantic word verification- to evaluate the effects of a naming treatment that involved semantic cuing. Different neural activation patterns were noted, depending on the task performed. Therefore, it may also be worthwhile to incorporate multiple tasks that evaluate different components of processing and these tasks need not correspond to the treatment task.

Before closing this section we would like to point out that it is important for researchers to ascertain, prior to scanning, a participant's ability to understand and perform the scanner tasks. A simulated scanner can be used for this purpose and participants can practice performing scanner tasks prior to participating in actual scanning sessions. In addition to allowing participants to acclimate to the scanning environment, this practice may diminish errors due to factors unrelated to a participant's abilities with the tasks of interest and, in addition, may help to decrease reaction times (for individuals with unusually delayed response patterns).

## 7.0 The Issue of Variability

One important issue in the design of neuroimaging studies with aphasic individuals is the possible variability in the measurements of neural activation over time that are not due to the treatment. This variability may be due to such things as scan-rescan variability, changes due to the natural history of a deficit, fluctuations in task performance and/or effects of task habituation and stimulus exposure.

A key challenge is to distinguish changes in pre- to post-treatment neural activity that are due to scan-rescan variability from changes due to the therapeutic intervention. Several

studies have addressed this particular issue by performing multiple pre-treatment (baseline) and post-treatment fMRI scans to develop an estimate of the variability in activation across repeated scans (Fridriksson, Morrow-Odom, Moser, et al., 2006; Fridriksson, Moser, Bonilha, et al., 2007). Fridriksson et al. (2006), for example, administered three fMRI scans prior to and after treatment, with aphasic individuals performing the same overt picture naming tasks during all sessions. Activation seen during baseline scans was compared with that seen during post-treatment scanning. This provided a powerful means for identifying activation changes from pre- to post-treatment that were reliable relative to the variability observed within the multiple pre and post scanning sessions.

The main drawback of this approach is that it is costly and labor intensive. Given this, another approach is to include control items in the fMRI task that are used to "index" variability not due to treatment. For example, Menke et al. (2009) examined naming performance on a large set of stimuli (object pictures) prior to treatment and scanning and selected items that the study participants could not name consistently for inclusion in the scan task. Half of the items were then trained, whereas the other half were untrained and brain activation was assessed prior to and following treatment for both sets. To control for any changes in activation across assessments associated with extraneous variables (i.e., variables other than the treatment provided), activation changes for the untreated items was removed from changes seen for the treated items using an exclusive mask. The rationale was that the changes in neural responses to these items at pre- and post-treatment provide a measure of repeated exposure and scan-rescan variability. This procedure provided an internal measure of reliability in the same participants, on the same task and at the same time allowed training effects to be evaluated. However, one issue related to the use of untrained items to index scan-rescan reliability is that it is a highly conservative measure of the effects of training as it assumes no generalization from trained to untrained items, something which is not always the case. Indeed, generalization from trained to untrained items is a common (and often desirable) effect of aphasia treatment. In fMRI treatment studies in which there is behavioral evidence of generalization, a possible approach to indexing variability would be use BOLD signal changes from just those untrained items that show no evidence of treatment generalization.

Another approach to dealing with variability is to use a control task to index changes unrelated to treatment. For example, Leger et al. (2002) administered two different fMRI tasks to an aphasic individual prior to and after a 6-week period of speech and language treatment: (a) overt naming, which evaluated an impaired function, and (b) rhyming, which evaluated an unimpaired function and was therefore considered a control condition. For the aphasic individual, results showed stable activation across imaging sessions for the rhyme task that was similar to that observed in six healthy participants. However, activation shifts were noted from pre- to post-treatment scans for the overt naming task, reflecting improved naming performance. Given the stability related to the rhyming task, the changes observed for the naming task could, with some confidence, be interpreted as reflecting recovery of some aspect of naming function rather than general language or cognitive changes. The challenge for this type of approach is in selecting a control task that is sufficiently distinct from the targeted function that treatment generalization is not expected, but not so different that it has no relationship to the functions of interest.

Finally, a related issue concerns the possibility of differences in longitudinal reliability for items depending on extent of exposure (expertise). Meltzer et al. (2009) assessed fMRI signal change magnitude in unimpaired volunteers across four scanning sessions at one-month intervals. Participants were asked to name pictures of the same item, new items, and explicitly over-learned items during each session, and activations across sessions were statistically compared for each type of item. Results indicated activation decreases across

scanning sessions, regardless of item type, but also revealed different degrees of between-session reliability for items of each type. The general decreases were attributed to task habituation, whereas, the reliability differences noted among stimulus types indicated that differential degrees of exposure (over-learning in particular) may affect the reliability of repeated fMRI measurements. These findings raise important questions for studies of aphasia recovery. Activation decreases from pre- to post-training could reflect task habituation and if responses are over-learned during aphasia treatment, this could impact the reliability of activation changes seen from pre- to post-treatment (see Meltzer et al. (2009) for further discussion of the implications of their findings for designing treatment studies in aphasia).

Distinguishing treatment changes (including generalization) from changes that are due to scan-rescan variability and other extraneous factors is a complex and difficult matter and approaches that involve the use of repeated scans and control items or tasks have specific strengths and weaknesses. Thus far we have focused on examples involving aphasia acquired as a consequence of cerebrovascular accident. As behavioral studies continue to reveal that treatment can have positive effects even in the context of progressive language disorders (e.g., Rapp & Glucroft, 2009), investigations of the neural changes that support these benefits will increase (Beeson, et al., 2011; Dressel, 2011; Marcotte & Ansaldo, 2010.) In this context, the specific challenge is to identify treatment-related neural changes against a backdrop of deterioration of language and possibly other cognitive functions. These investigations face the same types of challenges as have been discussed in this section.

## 8.0 Summary and Conclusions

An important objective of this paper is to discuss practices that will facilitate the interpretation of findings from neuroimaging investigations of aphasia treatment as well as the integration of findings across studies. Interpretation and integration are essential if we are to advance our quite limited understanding of the injured brain’s response to language treatment. To this end, we have identified and discussed key experimental design issues. Given the early stage in the development of neuroimaging of aphasia treatment research it is difficult to strongly advocate for specific best practices, nonetheless, we believe that it is important to identify critical challenges, and to alert researchers to issues of experimental design that they must consider in addressing these challenges.

In brief, we have made the following major points: (1) A clear characterization of the language deficits as well as the spared language and related cognitive functions is necessary and requires a comprehensive, theory-based evaluation. (2) Scanner tasks must be designed to allow for the identification of the neural substrates associated specifically with the treatment-based changes in language functions. Different approaches can be used successfully for this purpose, including, additive factors-based approaches contrasting critical factors in experimental and control tasks and the use of parametric designs or priming (neural habituation) paradigms. In all cases, interpretation of findings is facilitated by using tasks that produce reliable, interpretable results in neurologically intact participants. (3) A difficult question concerns whether the relevant substrates are best identified by using scanner tasks that are difficult or easy for participants, with each having advantages and disadvantages. Alternatives include designs that allow for contrasting correct and incorrect responses or the use of implicit tasks. (4) Distinguishing treatment-based neural changes from unwanted sources of variability -such as scan-rescan variability- is a major challenge for this type of research. While currently there is no single best method for dealing with this difficult problem, there are advantages of multiple pre and post-treatment scanning, and the use of untreated items or control tasks as indices of scan-rescan variability against which the treatment-based changes can be identified.

In the end, advances in research directed at understanding the neural substrates of recovery of language function will be determined, by the quality of "the two sides of the equation" – on one side, by the depth of our understanding of the affected language functions and, on the other, by our ability to design functional neuroimaging tasks that will reveal the neural activation patterns that specifically support the recovery of language functions.

## Acknowledgments

The authors wish to thank the School of Communication and the Vice President for Research at Northwestern University for providing funds to support the Neuroimaging in Aphasia Treatment Research Workshop held at Northwestern University, July 2009. Participants of the workshop, who contributed to the ideas in this paper include: Ana Ansaldo, Roelien Bastiaanse, Pelagie Beeson, Stephano Cappa, David Caplan, Leora Cherney, David Copland, Jenny Crinion, Bruce Crosson, Dirk den Ouden, Susan Edwards, Evy Fitch-Brink, Julius Fridriksson, Argye Hillis, Audrey Holland, Chien-Ju Hsu, Aneta Kielar, Monique King, Swathi Kiran, Jim Kloet, Sladjana Lukic, Marcus Meinzer, Charis Price, Ellyn Riley, Steve Small, Brenda Rapp, Dorothee Saur, Marion Smits, Yasmeen Faroqi-Shah, Cynthia Thompson, and Eisha Wali.

This work was supported by National Institutes of Health, Institute on Deafness and other Communication Disorders: R01DC006740 (BR), DC02146, DC003108 and DC011032 (DC) and R01DCOI948-17 and R01DC007213 (CT).

## References

Bachman DL, Morgan A. The role of pharmacotherapy in the treatment of aphasia: Preliminary results. Aphasiology. 1988; 2:225–228.

Beeson PM, King RM, Bonakdarpour B, Henry ML, Cho H, Rapcsak SZ. Positive Effects of Language Treatment for the Logopenic Variant of Primary Progressive Aphasia. Journal of Molecular Neuroscience. 201110.1007/s12031-011-9579-2

Bookheimer SY, Zeffiro TA, Blaxton T, Balow BA, Gaillard WD, Sato S, Kufta C, Fedio P, Theodore WH. A direct comparison of PET activation and electrocortical stimulation mapping for language localization. Neurology. 1997; 48 (4):1056–1065. [PubMed: 9109900]

Buckner RL, Andrews-Hanna JR, Schacter DL. The brain's default network. Anatomy, function, and relevance to disease. Annals of the New York Academy of Sciences. 200810.1196/annaks,1440.001

Byng S, Duchan JF. Social model philosophies and principles: Their applications to therapies for aphasia. Aphasiology. 2005; 19:906–922.

Cao Y, Vikingstad EM, George KP, Johnson AF, Welch KMA. Cortical language activation in stroke patients recovering from aphasia with functional MRI. Stroke. 1999; 30:2331–2340. [PubMed: 10548667]

Caplan D. Experimental design and interpretation of functional neuroimaging studies of cognitive processes. Human Brain Mapping. 2009; 30:59–77. [PubMed: 17979122]

Crosson B, McGregor D, Gopinath KS, Conway TW, Benjamin B, Chang Y-L, Moore AB, Raymer AM, Briggs RW, Sherod MT, Wierenga CT, White KD. Functional MRI of language in aphasia: A review of the literature and the methodological challenges. Neuropsychology Review. 2007; 17:157–177. [PubMed: 17525865]

Crinion JT, Leff AP. Recovery and treatment of aphasia after stroke: functional imaging studies. Current Opinion in Neurology. 2007; 20 (6):667–673. [PubMed: 17992087]

Dehaene S, et al. Cerebral mechanisms of word masking and unconscious repetition priming. Nat Neurosci. 2001; 4:752–758. [PubMed: 11426233]

Dressel K. Model-oriented naming therapy in semantic dementia: a single-case fMRI study. Aphasiology. 2011; 24:1–23.

Fedorenko E, Hsieh PJ, Nieto-Castanon A, Whitfield-Gabrieli S, Kanwisher N. New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. Journal of Neurophysiology. 2010; 104(2):1177–1194. [PubMed: 20410363]

Fridriksson J, Morrow-Odom L, Moser D, et al. Neural recruitment associated with anomia treatment in aphasia. Neuro Image. 2006; 32:1403–1412. [PubMed: 16766207]

Fridriksson J, Moser D, Bonilha L, Morrow KL, Shaw H, Baylis GC, Rorden C. Neural correlates of phonological and semantic-based anomia treatment in aphasia. Neuropsychologia. 2007; 45(8): 1812–22. [PubMed: 17292928]

Friston K, Price C, Fletcher P, Moore C, Frackowiak R, Dolan R. The trouble with cognitive subtraction. Neuro Image. 1996; 4:97–104. [PubMed: 9345501]

Kiran S, Thompson CK. Effects of exemplar typicality on naming in aphasia. Journal of Speech, Language, and Hearing Research. 2003; 46:608–822.

Leger A, Demonet JF, Ruff S, et al. Neural substrates of spoken language rehabilitation in an aphasic patient: an fMRI study. Neuroimage. 2002; 17(1):174–83. [PubMed: 12482075]

Marcotte K, Ansaldo A. The neural correlates of semantic feature analysis in chronic aphasia: discordant patterns according to the etiology. Semin Speech Lang. 2010; 31(1):52–63. [PubMed: 20221954]

Martin, N.; Thompson, CK.; Worrall, L., editors. Aphasia Rehabilitation: The Impairment and its Consequences. San Diego, CA: Plural Publishing Company; 2008.

Martin PI, Naeser MA, Ho M, Doron KW, Kurland J, Kaplan J, Wang Y, Nicholas M, Baker EH, Alonso M, Fregni F, Pascual-Leone A. Overt naming fMRI pre- and post-TMS: Two nonfluent aphasia patients, with and without improved naming post-TMS. Brain and Language. 2009; 111:20–35. [PubMed: 19695692]

McCandliss BD, Cohen L, Dehaene S. The visual word form area: expertise for reading in the fusiform gyrus. Trends in Cognitive Sciences. 2003; 7:293–299. [PubMed: 12860187]

Meinzer M, Flaisch T, Obleser J, Assadollahi R, Djundja D, Barthel G, et al. Brain regions essential for improved lexical access in an aged aphasic patient: A case report. BMC Neurology. 2006; 6:28. [PubMed: 16916464]

Meinzer M, Harnish S, Condway T, Crosson B. Recent developments in functional and structural imaging of aphasia recovery after stroke. Aphasiology. 2011; 25 (3):271–290. [PubMed: 21532927]

Meinzer M, Mohammadi S, Kugel H, Schiffbauer H, Flöel A, Albers J, et al. Integrity of the hippocampus and surrounding white matter is correlated with language training success in aphasia. Neuroimage. 2010; 53:283–290. [PubMed: 20541018]

Meltzer JA, Postman-Caucheteux WA, McArdle JJ, Braun AR. Strategies for longitudinal neuroimaging studies of overt language production. Neuro Image. 2009; 47 (2):745–755. [PubMed: 19427907]

Menke R, Meinzer M, Kugel H, Deppe M, Baumgartner A, et al. Imaging short- and long-term training success in chronic aphasia. BMC Neuroscience. 2009; 10:118. [PubMed: 19772660]

Mummery CJ, Ashburner J, Scott SK, Wise RJ. Functional neuroimaging of speech perception in six normal and two aphasic subjects. Journal of the Acoustical Society of America. 1999; 106:449–457. [PubMed: 10420635]

Naeser M, Martin P, Nicholas M, Baker E, et al. Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. Brain and Language. 2005; 93:95–105. [PubMed: 15766771]

Ohyama M, Senda M, Kitamura S, Ishii K, Mishina M, Terashi A. Role of the nondominant hemisphere and undamaged area during word repetition in poststroke aphasics. Stroke. 1996; 27:897–903. [PubMed: 8623110]

Peck K, Moore A, Crosson B, et al. Functional magnetic resonance imaging before and after aphasia therapy. Stroke. 2004; 35:554–559. [PubMed: 14739418]

Poldrack RA. Imaging brain plasticity: Conceptual and methodological issues – a theoretical review. Neuroimage. 2000; 12:1–13. [PubMed: 10875897]

Postman-Caucheteux WA, Birn RM, Pursley RH, Butman JA, Solomon JM, Picchioni D, McArdle J, Braun AR. Single-trial fMRI Shows Contralesional Activity Linked to Overt Naming Errors in Chronic Aphasic Patients. J Cogn Neurosci. 2010; 22:1299–1318. [PubMed: 19413476]

Price C, Crinion J. The latest on functional imaging studies of aphasic stroke. Current Opinion in Neurology. 2005; 18 (4):429–434. [PubMed: 16003120]

Price CJ, Friston KJ. Functional imaging studies of neuropsychological patients: Applications and limitations. Neurocase. 2002; 8:345–354. [PubMed: 12499409]

Rapp B, Glucroft B. The benefits and protective effects of behavioral treatment for dysgraphia in a case of Primary Progressive Aphasia. Aphasiology. 2009; 23:236–265. [PubMed: 21603153]

Rapp B, Goldrick M. Speaking words: Contributions of cognitive neuropsychological research. Cognitive Neuropsychology. 2006; 23 (1):39–73. [PubMed: 21049321]

Rapp B, Vindiola M. The neural consequences of behavioral intervention in dysgraphia. Brain and Language. 2005; 95:237–238.

Santi A, Grodzinsky Y. Taxing working memory with syntax: Bihemispheric modulations. Human Brain Mapping. 2007; 28:1089–1097. [PubMed: 17133392]

Sebastian R, Kiran S. Task Modulated activation patterns in chronic stroke patients with aphasia. Aphasiology. 2011; 25(8):927–951.

Talairach, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain. New York: Thieme; 1988.

Thompson, CK. Plasticity of language networks. In: Baudry, M.; Bi, X.; Schrieber, SS., editors. Syntaptic plasticity: Basic mechanisms to clinical applications. New York: Marcel Dekker, Inc; 2005. p. 343-355.

Thompson CK, Bonaakdarpour B, Fix S. Neural substrates of verb argument structure processing in agrammatic aphasic and healthy age-matched listeners. Journal of Cognitive Neuroscience. 2010; 22(9):1993–2011. [PubMed: 19702460]

Thompson CK, Den Ouden DB. Neuroimaging and recovery of language in aphasia. Current Reports in Neurology and Neuroscience. 2008; 8:475–483.

Thompson CK, Den Ouden DB, Bonakdarpour B, Garibaldi K, Parrish TB. Neural plasticity and treatment-induced recovery of sentence processing in agrammatism. Neuropsychologia. 2010 epub ahead of print.

Tsapkini K, Vindiola M, Rapp B. Patterns of brain reorganization subsequent to left fusiform damage: fMRI evidence from visual processing of words and pseudowords, faces and objects. Neuro Image. 2011; 55:1357–1372. [PubMed: 21168516]

Vitali P, Abutalebi J, Tettamanti M, Danna M, Ansaldo AI, Perani D, et al. Training-induced brain remapping in chronic aphasia: A pilot study. Neurorehabilitation and Neural Repair. 2007; 21:152–160. [PubMed: 17312090]

## Highlights

- Experimental design issues in functional neuroimaging research on aphasia treatment
- Using fMRI to identify neural changes supporting treatment and recovery in aphasia
- Relating treatment-based changes in language functions to associated neural changes

## Table 1

Language profile considered by Menke et al (2009) to be characteristic of a specific deficit to the linking of semantic and spoke word form representations.

| Language Task | Performance |
|---|---|
| Hearing | Normal |
| Auditory comprehension | Adequate |
| Object naming | |
| Accuracy | Poor |
| Error types | Semantic |
| Responsiveness to phonological cues | Yes |
| Word fluency | Severely reduced |
| Repetition | Adequate |
| Apraxia of speech | Mild, if present |