Do women really have more bilateral language representation than men? A meta-analysis of functional imaging studies

Iris E. C. Sommer¹, André Aleman¹, Anke Bouma² and René S. Kahn¹

¹Department of Psychiatry, University Medical Center, Utrecht, The Netherlands and ²Department of Psychology, University of Groningen, The Netherlands

Summary

Sex differences in cognition are consistently reported, men excelling in most visuospatial tasks and women in certain verbal tasks. It has been hypothesized that these sex differences in cognition result from a more bilateral pattern of language representation in women than in men. This bilateral pattern of language representation in women is thought to interfere with visuospatial functions in the right hemisphere. To test whether language representation is indeed more bilateral in the female than in the male brain, a meta-analysis was performed on studies that assessed language activity with functional imaging in healthy men and women. Effect sizes were weighted for sample size and the meta-analytic method was applied to obtain a combined effect size. Fourteen studies were included, providing data on 377 men and 442 women. Meta-analysis yielded a mean weighted effect of 0.21 with a 95% confidence interval of −0.05 to 0.48, indicating no significant difference in language lateralization between men and women. This implies that the putative sex difference in language lateralization may be absent at the population level, or may be observed only with some, as yet not defined, language tasks. It is therefore not likely that differences in language lateralization underlie the general sex differences in cognitive performance, and the neuronal basis for these cognitive sex differences remains elusive.

Keywords: language lateralization; sex difference; functional imaging; meta-analysis


Introduction

Sex differences in cognitive performance are consistently reported, men excelling in mental rotation and spatial perception and women performing better on verbal memory tasks, verbal fluency tasks and in speed of articulation (Linn and Petersen, 1985; Kimura, 2000). Furthermore, language and reading disorders are reported to occur approximately twice as often in boys than in girls (Flannery et al., 2000). Since the sex difference in the risk of language disorders may be associated with the sex difference in cognitive skills, it would be of both scientific and clinical importance to obtain insight into the mechanism that underlies these sex differences in cognitive functions.

The cerebral substrate of the sex differences in cognition is not known. It has been hypothesized that language functions are represented more bilaterally in the female brain than in the male brain (McGlone, 1980; Dorion et al., 2000; Gur et al., 2000). Women might thus use both hemispheres for language functions while males predominantly use the left hemisphere. A more bilateral pattern of language representation is thought to result in better verbal skills, while visuospatial processing would be inferior in subjects with bilateral language representation (Levy, 1969). Thus, the female deficit in spatial performance is hypothesized to result from competition between verbal and spatial functions in the right hemisphere.

The theory that sex differences in cognition arise from more bilateral representation of language functions in females than in males is supported by two findings. First, female stroke patients have been reported to exhibit verbal impairment less frequently after lesions of the left hemisphere than male patients (McGlone, 1980). Secondly, structural MRI studies demonstrated that asymmetry of the planum temporale (the upper surface of the temporal lobe largely overlapping with
Several studies could not be included in the meta-analysis because sufficient statistical data were not available. To yield a global impression of the mean findings of these studies, a vote-count analysis was also carried out. For this test, each study was given a weight, based on its sample size. The total weights of all studies that reported a sex difference in language lateralization (more asymmetry in men than in women) was compared with the total weights of all studies that reported no sex difference (cf. Sommer et al., 2003a). All studies were included in this vote-count analysis to obtain a general picture of all published evidence. In studies that applied two or three language tasks, scan data of all tasks were used (i.e., these subjects were counted twice or three times).

Results
Twenty-four studies had been selected that had measured language lateralization with functional imaging techniques in healthy men and women and reported on a possible sex difference. Characteristics and main findings of the included studies are provided in Table 1.

Meta-analysis
From these studies, fourteen could be included in the meta-analysis, providing data on 377 men and 442 women. Figure 1 shows the individual weighted effect size of each study. The meta-analysis yielded the following results: mean weighted effect: \( d = 0.21 \), 95% confidence interval \(-0.04 \) to 0.32). This indicates no significant difference in language lateralization between men and women. The homogeneity index \( Q \) was 31.8, \( P = 0.003 \), indicating significant heterogeneity among the studies. A moderator analysis was performed with the variable ‘language activation task’; this showed no significant difference between word production tasks (i.e., verbal fluency and verb generation): \( k = 5, n = 491, d = 0.14 \) (–0.04 to 0.32) and receptive language tasks (semantic decision tasks): \( k = 6, n = 254, d = 0.05 \) (–0.20 to 0.31), \( Q(b) = 0.32, P = 0.57 \). There were too few studies with other language tasks to investigate possible differences between other task characteristics, such as word versus non-word, or single words versus sentences.

Vote count analysis
All 24 studies were included in the vote-count analysis, providing data on 619 men and 743 women. From the 24 studies, 15 studies reported having found no sex difference in language lateralization. Nine studies reported finding a sex difference for at least one language task. The vote-count revealed a score of 1137 for studies that found no sex difference, compared with 285 for studies that did find a sex difference. This implies that the studies that did report a sex difference generally had a much smaller sample size (mean 31, SD 10) than the studies that reported no difference between the sexes (mean 76, SD 24).

Discussion
In this study, data on language lateralization of healthy men and women from 14 functional imaging studies were
combined in a meta-analysis. The difference in language lateralization between men and women was not significant. The vote-count analysis, which included 24 studies, showed that the majority of studies reported no difference in lateralization between the sexes. Studies that did report a sex difference in language lateralization had smaller sample sizes than studies that reported no sex difference.

Three hypotheses may be considered in the light of these findings. First, there may be a sex difference at the population level, but it is relatively small so that it is only sporadically observed. Were this to be true, studies with larger sample sizes would be expected to report a sex difference in lateralization more frequently than studies with smaller sample sizes, since they have more power to detect subtle differences. However, the vote-count analysis revealed that the studies that reported a sex difference had much smaller sample sizes than the studies with negative findings. Thus, the hypothesis of a true but subtle sex difference in language lateralization at the population level is not supported by our data.

A second hypothesis to explain the present findings is that sex differences in language lateralization may be task-dependent. Indeed, there was significant heterogeneity among the studies in our meta-analysis, which may be congruent with this hypothesis. Shaywitz and colleagues stated that only tasks requiring phonological processing yield more bilateral language lateralization in women, while semantically processed tasks yield no sex difference in lateralization (Shaywitz et al., 1995). The results of the studies listed in Table 1 do not support this statement, since sex differences in lateralization have been reported with several semantic language tasks (semantic decision tasks, verbal fluency, verb generation and story listening). Alternatively, Kansaku

### Table 1. Functional imaging studies (fMRI, functional transcranial Doppler or PET) on sex differences in language lateralization

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of healthy subjects</th>
<th>Language task</th>
<th>Authors report sex difference in lateralization</th>
<th>Effect size or reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binder et al., 1995</td>
<td>2 men, 3 women</td>
<td>Semantic decision</td>
<td>No</td>
<td>Included, $d = 0.17$</td>
</tr>
<tr>
<td>Buckner et al., 2001</td>
<td>12 men, 20 women</td>
<td>Stem completion</td>
<td>No</td>
<td>Excluded, insufficient data</td>
</tr>
<tr>
<td>Shaywitz et al., 1995</td>
<td>19 men, 19 women</td>
<td>Rhyme judgement</td>
<td>Yes</td>
<td>Included, $d = 1.22$</td>
</tr>
<tr>
<td>Pugh et al., 1996</td>
<td>19 men, 19 women</td>
<td>Rhyme judgement</td>
<td>Yes</td>
<td>Excluded, same subjects and scans as Shaywitz, 1999</td>
</tr>
<tr>
<td>Jaeger et al., 1998</td>
<td>9 men, 8 women</td>
<td>Past tense generation</td>
<td>Yes</td>
<td>Excluded: insufficient data</td>
</tr>
<tr>
<td>Van der Kallen et al., 1998</td>
<td>10 men, 10 women</td>
<td>Verbal fluency</td>
<td>No</td>
<td>Included, $d = 0.42$</td>
</tr>
<tr>
<td>Schlosser et al., 1998</td>
<td>6 men, 6 women</td>
<td>Verbal fluency</td>
<td>Yes</td>
<td>Excluded, insufficient data</td>
</tr>
<tr>
<td>Springer et al., 1999</td>
<td>5 men, 4 women</td>
<td>Verbal generation</td>
<td>No</td>
<td>Included, $d = 0.9$</td>
</tr>
<tr>
<td>Frost et al., 2000</td>
<td>52 men, 48 women</td>
<td>Semantic decision</td>
<td>No</td>
<td>Included, $d = 1.5$</td>
</tr>
<tr>
<td>Pujol et al., 1999</td>
<td>50 men, 50 women</td>
<td>Verbal fluency</td>
<td>No</td>
<td>Included, $d = 0.1$</td>
</tr>
<tr>
<td>Vingerhoets et al., 1999</td>
<td>38 men, 52 women</td>
<td>Several verbal tasks</td>
<td>No</td>
<td>Excluded: no separate data for spatial and verbal tasks</td>
</tr>
<tr>
<td>Gur et al., 2000</td>
<td>14 men, 13 women</td>
<td>Semantic decision</td>
<td>Yes</td>
<td>Included, $d = 0.03$</td>
</tr>
<tr>
<td>Kansaku et al., 2000</td>
<td>16 men, 14 women</td>
<td>Story listening</td>
<td>Yes</td>
<td>Included, $d = 1.01$</td>
</tr>
<tr>
<td>Knecht et al., 2000a</td>
<td>77 men, 111 women</td>
<td>Verbal fluency</td>
<td>No</td>
<td>Excluded, subjects overlap with Knecht et al., 2000b</td>
</tr>
<tr>
<td>Knecht et al., 2000b</td>
<td>128 men, 198 women</td>
<td>Verbal fluency</td>
<td>No</td>
<td>Included, $d = 0.14$</td>
</tr>
<tr>
<td>Philips et al., 2001</td>
<td>10 men, 10 women</td>
<td>Story listening</td>
<td>Yes</td>
<td>Included, $d = 1.76$</td>
</tr>
<tr>
<td>Vikingstad et al., 2000</td>
<td>17 men, 19 women</td>
<td>Verbal generation</td>
<td>Yes</td>
<td>Included, $d = 0.01$</td>
</tr>
<tr>
<td>Billingsley et al., 2001</td>
<td>6 men, 5 women</td>
<td>Rhyme task</td>
<td>No</td>
<td>Excluded: insufficient data</td>
</tr>
<tr>
<td>Sommer et al. 2003b</td>
<td>12 men, 12 women</td>
<td>Semantic decision</td>
<td>No</td>
<td>Included, $d = -0.76$</td>
</tr>
<tr>
<td>Baxter et al., 2003</td>
<td>9 men, 10 women</td>
<td>Semantic decision</td>
<td>Yes</td>
<td>Excluded: insufficient data</td>
</tr>
<tr>
<td>Hund-Georgiadis et al., 2002</td>
<td>18 men, 16 women</td>
<td>Semantic decision</td>
<td>No</td>
<td>Included, $d = -0.36$</td>
</tr>
<tr>
<td>Rossell et al., 2002</td>
<td>6 men, 6 women</td>
<td>Target detection of words</td>
<td>Yes</td>
<td>Excluded: hemifield projection of stimuli</td>
</tr>
<tr>
<td>Szafiarski et al., 2002</td>
<td>24 men, 26 women</td>
<td>Semantic decision</td>
<td>No</td>
<td>Included, $d = 0.07$</td>
</tr>
<tr>
<td>Sommer et al., 2004</td>
<td>10 men, 14 women</td>
<td>Verb generation</td>
<td>No</td>
<td>Excluded: insufficient data</td>
</tr>
</tbody>
</table>
and Kitazawa (2001) suggested that only tasks that present real words instead of non-words elicit a sex difference. This suggestion could not be supported either, since two studies reported a sex difference in lateralization with rhyme tasks consisting of non-words (Shaywitz et al., 1995; Pugh et al., 1996).

Furthermore, inspection of the data listed in Table 1 revealed no clusters of positive findings with a certain type of language task. The moderator analysis on task characteristics found no significant difference between productive and receptive language tasks. Thus, we found no support for this hypothesis.

The third hypothesis is the null hypothesis: that there is no sex difference in language lateralization at the population level. If this hypothesis were true, the sex differences reported in the studies with small samples may reflect biased reporting of chance findings, i.e. the ‘file drawer’ problem (Rosenthal, 1991). This hypothesis offers an explanation for the larger mean sample size of studies with negative findings compared with those with positive findings. Since the first two hypotheses appeared unlikely, the third hypothesis would best accommodate the present findings.

Functional imaging studies on sex differences in lateralization have not been reviewed previously with a meta-analytic technique. However, sex differences in language lateralization have been estimated with other methods, such as clinical studies on patients with unilateral brain lesions and experiments in which dichotic listening tests are applied to measure perceptual asymmetry.

Current knowledge on the organization of the brain for language in both sexes is largely based on data from studies that establish a link between language functions and brain organization by associating disrupted function of a brain area with a change in linguistic behaviour, usually a deficit. Such studies identify a brain area as critical for a certain aspect of language, which means that aphasia results when a critical area is damaged. This type of study, which includes frequency studies of different kinds of aphasia and their corresponding lesions, observations using the carotid sodium Amytal (Wada) procedure and findings with intra-operative electrical stimulation, yields information about the lateralization of critical language areas, such as Brodmann’s Area (BA) 44 and 45 and Wernicke’s area (the upper part of the superior temporal gyrus). In contrast, the studies that were included in this meta-analysis record physiological measures of brain activity while subjects were engaged in tasks that addressed certain language functions, such as PET, functional MRI and functional transcranial Doppler. Language dominance as measured with these techniques is highly correlated to dominance as assessed with the Wada procedure, which is considered the gold standard (Binder et al., 1997; Deppe et al., 2000). However, the degree of language lateralization

<table>
<thead>
<tr>
<th>Citation</th>
<th>Effect</th>
<th>NTotal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binder 1995</td>
<td>-.124</td>
<td>5</td>
</tr>
<tr>
<td>Gur 2000</td>
<td>-.029</td>
<td>27</td>
</tr>
<tr>
<td>Hund-Georgiadis 2002</td>
<td>-.351</td>
<td>34</td>
</tr>
<tr>
<td>Kansaku 2000</td>
<td>.983</td>
<td>30</td>
</tr>
<tr>
<td>Knecht 2000</td>
<td>.140</td>
<td>326</td>
</tr>
<tr>
<td>Phillips 2000</td>
<td>1.686</td>
<td>20</td>
</tr>
<tr>
<td>Pujol 1999</td>
<td>.099</td>
<td>100</td>
</tr>
<tr>
<td>Shaywitz 1995</td>
<td>1.194</td>
<td>38</td>
</tr>
<tr>
<td>Sommer 2001</td>
<td>-.734</td>
<td>24</td>
</tr>
<tr>
<td>Springer 1999</td>
<td>-.149</td>
<td>100</td>
</tr>
<tr>
<td>Szaflarski 2002</td>
<td>.069</td>
<td>50</td>
</tr>
<tr>
<td>Van de Kallen 1998</td>
<td>.402</td>
<td>20</td>
</tr>
<tr>
<td>Vikingstad 2000</td>
<td>.010</td>
<td>36</td>
</tr>
<tr>
<td>Xiong 1998</td>
<td>.800</td>
<td>9</td>
</tr>
<tr>
<td>Random Combined (14)</td>
<td>.214</td>
<td>819</td>
</tr>
</tbody>
</table>

**Fig. 1** Language processing in men and women.
measured with functional imaging techniques may be lower than that observed in lesion studies, since language activation may also be detected at sites that are not critical for that language function, but may be activated for non-specific supporting functions. Examples of areas that are frequently activated during language tasks with functional imaging, but that do not produce aphasia when lesioned, are the anterior cingulate gyrus and the superior frontal gyrus (Binder et al., 1997).

McGlone (1980) first reported a higher incidence of aphasia in 23 men compared with 20 women after unilateral lesions of the left hemisphere. Kimura (1983) replicated McGlone’s finding of a higher incidence of aphasia after left-hemisphere injury in a sample of 144 men and 92 women. However, if this sex difference in the incidence of aphasia were the result of a more bilateral pattern of language representation in women than in men, the incidence of aphasia after right-hemisphere injury would be expected to be higher in women than in men. This prediction was not met, since the incidence of aphasia was similar (2% for men and 1% for women) in a group of 134 men and 100 women with lesions of the right hemisphere (Kimura, 1983). However, Kimura (1983) found that aphasia in women was more common after anterior lesions than after posterior lesions within the left hemisphere, whereas in men aphasia was more common after lesions of the posterior language areas (Kimura, 1983). Because anterior lesions occur less frequently than lesions of the posterior language areas, this offers an alternative explanation for the lower incidence of aphasia in women after left-sided lesions. McGlone’s and Kimura’s finding of higher frequencies of aphasia in men than in women after left-hemisphere lesions was contrasted by Kertesz and Sheppard (1981), who reported that 78 right-handed females who had suffered left hemisphere stroke performed slightly worse on the Western Aphasia Battery (WAB) than 114 males with similar lesions. No difference in score on the WAB between men and women was found after right-hemisphere stroke. Thus, a sex difference after left-hemisphere lesions has not been consistently reported and differences in aphasia after right-hemisphere lesions have never been demonstrated.

Another test of sex differences in cerebral dominance for language is provided by experimental studies using dichotic listening techniques. Dichotic listening studies compare the performance between items presented to the left and to the right ear. If this perceptual asymmetry is reduced, this is taken to reflect decreased cerebral dominance. However, decreased language dominance is not the only factor that can cause low perceptual asymmetry, since this method is also liable to differences in inhibition, selected attention and sustained concentration. Furthermore, results of dichotic listening studies are largely dependent on the choice of task that is applied. Sex differences in dichotic listening tests have been reported by several studies (Lake and Bryden, 1976; Voyer, 1996; Coney, 2002), but not by all (Witelson, 1976; Carter-Salzman, 1979; Demarest and Demarest, 1981). A few studies reported larger perceptual asymmetry in women than in men (van Duyn and Sass, 1979; Hiscock and Hiscock, 1988). Furthermore, two large dichotic listening studies reported no gender difference in language lateralization. Hiscock and MacKay (1985) administered verbal dichotic listening tests to 477 right-handed volunteers in five consecutive experiments. None of the five analyses yielded a significant sex difference, and even when data of the five tests were pooled no sex difference prevailed. Similarly, Hugdahl (2003) found no sex difference in a database of 1018 healthy subjects performing a verbal dichotic listening test.

Some authors suggest that language dominance is less stable in women, since it is thought to fluctuate during the menstrual cycle (Altemus et al., 1989). However, this hypothesized hormonal effect has only been studied using perceptual asymmetry, with contradicting findings (Altemus et al., 1989; Sanders and Wenmoth, 1998; Hausmann et al., 2002). Therefore, the effects on perceptual asymmetry may be the result of differences in performance rather than true differences in cerebral dominance. Thus, the results from dichotic listening studies are inconsistent. Since two large studies found no sex difference in language lateralization, the sex difference, if present, is supposedly subtle. These findings are in line with the result of this meta-analysis.

In parallel to the findings on functional language lateralization, studies that measure the anatomical size of the planum temporale are also inconsistent on a gender difference in asymmetry. Several studies reported that asymmetry of the planum temporale was larger in men (Wada et al., 1975, Bilder et al., 1994, Kulynych et al., 1994; Foundas et al., 2002), which probably results from a larger left planum temporale in men compared with women (Kulynych et al., 1994; Foundas et al., 2002). On the other hand, a number of studies could not observe a sex difference in asymmetry of the planum temporale (Kertesz et al., 1986; Duara et al., 1991; DeLisi et al., 1994; Jancke et al., 1994; Petty et al., 1995). A large voxel-based analysis on MRI scans of 465 healthy adults reported increased asymmetry of the planum temporale in men compared with women, which was caused by smaller left planum in women (Good et al., 2001). However, there are many studies on sex differences in structural asymmetry and they have not been meta-analysed yet, which makes it difficult to come to a conclusion on this topic. Furthermore, it is not clear whether greater asymmetry of the planum temporale in men than in women reflects a higher degree of functional language lateralization in men. Two studies investigated the correlation between asymmetry of the planum temporale and cerebral dominance for language. Foundas and colleagues reported that asymmetry of the planum temporale correlated to cerebral dominance assessed with the Wada test in 11 subjects (Foundas et al., 1994), while Tzourio and colleagues found no correlation between the same measurements in 14 subjects (Tzourio et al., 1998). The latter study did find a correlation between cerebral dominance and the size of the left planum. Thus, it is currently not clear whether
asymmetry of the planum temporale is an accurate predictor of functional language representation.

This short overview of the literature, together with the present findings, suggests that there may not be a sex difference in cerebral dominance for language. Future research may be aimed at other cerebral systems that could underlie sex differences in cognition. Several alternatives have been proposed. Kimura (2000) observed that women rely more on the anterior language area of the left hemisphere for language functions, while men mainly activate the posterior (temporoparietal) language area. It would be interesting to investigate if functional imaging studies can replicate her finding.

Differences in the width and shape of the corpus callosum have also been proposed to underlie cognitive sex differences. DeLacoste-Utamsing and Holloway (1982) first reported that the posterior part of the corpus callosum was larger and more bulbous in women. Since this first report there have been both confirmations and denials of sex differences in the width of the corpus callosum. In a review, Driesen and Raz (1995) concluded that there probably is a small difference favouring women. In addition, other commissures of the brain have also been reported to be wider in women than in men. The cross-sectional area of the anterior commissure is larger in women than in men (Allen and Gorski, 1991). Furthermore, the massa intermedia, which connects the two halves of the thalamus, is found to be more frequently absent in men than in women (Allen and Gorski, 1991). Possibly, the two hemispheres of the female brain may be better connected than the hemispheres of the male brain, which may give a higher speed of information transfer between the hemispheres in females, and could explain the female advantages in some language tasks.

Another alternative cerebral substrate for the sex differences in cognition has been described by Witelson (1995), who studied cytoarchitecture in post-mortem brains. In a small sample she found that the density of neurons in layers II and IV of the posterior temporal cortex was greater by 11% in women, with no overlap of scores between the sexes. However, this study has not been replicated.

This study has several limitations. First, a relatively large number of studies could not be included in the meta-analysis since they provided only qualitative information or insufficient statistical data about sex differences in lateralization. However, the sample size of the meta-analysis is still relatively large and the result of the vote-count analysis is congruent with the finding of the meta-analysis. Another limitation is that we analysed all language activation tasks that are generally lateralized to the left hemisphere, without dividing them into specific categories. There are many possible subdivisions that could be made on the basis of task characteristics, such as visual or auditory presentation, single-word stimuli or stories, phonological or semantic processing, productive or receptive tasks and abstract or concrete words. All these divisions appear to be equally valid candidates that may affect the degree of lateralization differently in the two sexes. However, all these separate categories would decrease the sample size, decrease the power of the meta-analysis and increase the chance of a false-positive finding. Therefore, we preferred to analyse the whole sample of all left-lateralized language activation tasks. We did, however, perform a moderator analysis, which showed that there was no difference in sex effect between productive and receptive language tasks.

In summary, this meta-analysis found no significant sex difference in functional language lateralization in a large sample of 377 men and 442 women. Thus, the hypothesis that language functions are generally presented more bilaterally in women than in men is not supported. This suggests that language lateralization is unlikely to underlie sex differences in cognition, and their biological basis remains elusive.

References


Dorion AA, Chantome M, Hasboun D, Zouaoui A, Marsault C, Capron C, Duyne M. Hemispheric asymmetry and corpus callosum


Lipsey MW and Wilson DB. The way in which intervention studies have ‘personality’ and why it is important to meta-analysis. Eval Health Prof 2001; 24: 236–54.


Pujol J, Deus J, Losilla JM, Capdevila A. Cerebral lateralization of language in normal left-handed people studied by functional MRI. Neurology 1999; 52: 1038–43.


