Evidence for bilateral involvement in idiom comprehension: An fMRI study

Monika-Zita Zempleni,a,b,c, Marco Haverkort,d Remco Renken,b,c and Laurie A. Stowea,b,c

Department of General Linguistics, University of Groningen, The Netherlands
School of Behavioral and Cognitive Neurosciences, University of Groningen, The Netherlands
Neuroimaging Center, University of Groningen, The Netherlands
Department of Linguistics, University of Nijmegen, The Netherlands

Received 31 January 2006; revised 22 September 2006; accepted 25 September 2006
Available online 1 December 2006

The goal of the current study was to identify the neural substrate of idiom comprehension using fMRI. Idioms are familiar, fixed expressions whose meaning is not dependent on the literal interpretation of the component words. We presented literally plausible idioms in a sentence forcing a figurative or a literal interpretation and contrasted them with sentences containing idioms for which no literal interpretation was available and with unambiguously literal sentences. The major finding of the current study is that figurative comprehension in the case of both ambiguous and unambiguous idioms is supported by bilateral inferior frontal gyri and left middle temporal gyrus. The right middle temporal gyrus is also involved, but seems to exclusively process the ambiguous idioms. Therefore, our data suggest a bilateral neural network underlying figurative comprehension, as opposed to the exclusive participation of the right hemisphere. The data also provide evidence against proposed models of idiom comprehension in which literal processing is by-passed, since figurative processing demanded more resources than literal processing in the language network.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Left hemisphere; Right hemisphere; Bilateral; Inferior frontal gyrus; Middle temporal gyrus; Idiom comprehension; Figurative language; Ambiguity resolution; Functional MRI

Introduction

This paper reports a functional magnetic resonance imaging (fMRI) study addressing the neural substrate of figurative language comprehension. The term, figurative language refers to metaphors and idioms, whose meaning is not dependent on the literal interpretation of their component words. Metaphors require the extension of the literal meaning to a more abstract representation (e.g., ‘that teacher is a real shark’ or ‘he is a star rising on the horizon’). Idioms can be regarded as ‘frozen’ metaphors, as they are figurative language but at the same time are familiar fixed multi-word expressions (e.g., ‘smell a rat’). Figurative elements are frequently used in natural languages perhaps because they allow speakers to express abstract concepts in a concise manner. In the current study we will focus on idiom processing. Idiom comprehension appears to depend on a complex neural network as evinced by the fact that it is impaired in several brain disorders, such as Alzheimer’s disease (Papagno et al., 2003; Papagno, 2001), schizophrenia (Titone et al., 2002), corpus callosum agenesis (Huber-Okrainec et al., 2005), right hemisphere damage and left hemisphere damage. Neuroimaging studies on figurative language have been limited to metaphor processing (Rapp et al., 2004; Bottini et al., 1994), so the goal of the current fMRI study was to evaluate the neural background of idiom comprehension.

Idioms form a heterogeneous group and can be classified according to several linguistic features. It is beyond the scope of the current paper to discuss definitions, typologies and terminologies of figurative language research in detail; for reviews see Glucksberg, 2001; Titone and Connine, 1999, 1994; Everaert et al., 1995, Nunberg et al., 1994; Cacciari and Tabossi, 1993. Here we will focus on those characteristics of idioms which are relevant for the current study.

First, as said above, idioms are by definition figurative language that is their meaning cannot be directly derived from the literal meanings of the constituent words. Instead, the overall figurative interpretation has to be constructed in light of a wider context. In the current study we will investigate the role of figurative versus literal processing in idiom comprehension by contrasting sentences with a figurative meaning versus sentences with a literal meaning. This contrast allows us to identify areas which are primarily involved in figurative comprehension. The figurative sentences contained an idiom within a sentence which supported an overall figurative interpretation. The literal sentences either did not contain an idiom at all, or contained a potential idiom, but only its literal interpretation was plausible (e.g., ‘Down
in the tunnel under the barn, the terrier smelled a rat.’). Therefore, the overall meaning of the sentence was literal.

Neuroimaging studies on the neural substrate of metaphor processing have shown increased activation for figurative comprehension when compared to a literal baseline (Rapp et al., 2004; Bottini et al., 1994). Although idioms are also figurative elements, they differ in the sense that they are fixed, i.e., ‘frozen’ expressions. There is considerable debate in the psycholinguistic literature as to whether idioms are processed like metaphors or rather processed as lexical items. According to Jackendoff (2002) idioms might represent a general tendency in language for lexicalization, i.e., the meaning and structure of word combinations which frequently co-occur might be stored as a sequence in the mental lexicon. According to the direct access models, e.g., the lexicalization hypothesis by Swinney and Cutler (1979), linguistic processing can be by-passed during idiom comprehension due to the storage of these expressions. If this is the case, we might expect a substantially different pattern for idioms than for metaphors, or in the most extreme version, even find more activation for the literal sentences than for the idiomatic ones. However, other authors (Gibbs, 1993) argue that idioms are not ‘dead’ metaphors since the underlying conceptual metaphors are always activated during idiom comprehension. In this case, the idiomatic sentences of the current study are expected to behave similarly to metaphors. There are also models in between these two extremes, suggesting that idioms vary in the degree of their metaphoricity (Nunberg et al., 1994; Glucksberg, 1993) with some being highly transparent (e.g., ‘smell a rat’) whereas others are non-transparent (e.g., ‘kick the bucket’) that is the metaphorical interpretation would be difficult to construct without knowing the idiom. In the current study, relatively transparent idioms were used (see Materials and methods section below); this factor was not manipulated.

Another variable which defines subclasses of idioms is literality (Titone and Connine, 1994): the extent to which idioms have an alternative plausible literal meaning, i.e., whether they are ambiguous between a literal and an idiomatic meaning. Idioms such as ‘kick the bucket’ or ‘smell a rat’ have literal interpretations along with the figurative ones. Others do not, e.g., ‘trip the light fantastic’ or ‘to be in a blue funk’. Ambiguity is a factor which may influence the processing of certain idioms, because a context-based choice of meaning is necessary when multiple meanings are available. In this sense, literally plausible idioms may be similar to lexical semantic ambiguities, i.e., words with multiple meanings such as ‘bank’, whose comprehension requires ambiguity resolution. In order to investigate the effects of ambiguity processing in idiom comprehension, we contrasted ambiguous and unambiguous sentences. As already stated above, we used literally plausible idioms in sentences which supported an overall literal interpretation of the sentence despite the presence of a potential idiom. An example using the analogue idiom in English cited above in a literal context would be ‘Down in the tunnel under the barn, the terrier smelled a rat’. The same idioms were also presented in sentences which were more consistent with the figurative interpretation. An analogue example in English would be ‘During the testimony of the witness, the jury smelled a rat’. These ambiguous sentences were contrasted with unambiguous sentences, which either contained an unambiguous, i.e., literally implausible, idiom (similar to the English sentence ‘Due to the strike of the railroad-workers, the travelers were in a blue funk’) or did not contain an idiom at all (for example ‘Due to the workload of the employees, the manager hired a new caseworker’), in order to create a fully factorial design. Note that the design was therefore not a within item but a mixed design.

To sum up, the current study used four types of experimental sentences: ambiguously idiomatic sentences (idiomatic context with an ambiguous idiom; aI); unambiguously idiomatic sentences (idiomatic context with unambiguous idiom; aL); ambiguously literal sentences (literal sentence with an ambiguous potential idiom; uI) and unambiguously literal sentences (uL). The Dutch experimental sentences used in the current study are illustrated in Table 1; here idioms are marked with bold typeface; in the case of the ambiguous conditions the biasing context is printed in italics.

Our primarily goal with this experiment was to identify the components of the neural network underlying idiom comprehension which, as we noted above, appears to be extensive enough to be easily disrupted by many brain disorders and lesion sites. Additionally, our goal with the current design was to separate figurative and ambiguity resolution components of idiom comprehension in order to identify which brain areas within this network are responsible for these different cognitive functions. We were also particularly interested in the lateral distribution of the network and the specific role of the two hemispheres since there is evidence in the literature, briefly discussed below, suggesting that the two hemispheres might be differentially sensitive to metaphoricity and/or ambiguity. Due to the sparse investigation of these functions to date, we refrain from formulating hypotheses as to which specific brain regions might be involved in this network, but we consider the potential role of the hemispheres in the following sections.

It has been shown that right hemisphere damage disrupts idiom comprehension (Kempler et al., 1999; Van Lancker and Kempler, 1987). In these studies a picture-matching auditory comprehension task was applied and the test materials consisted of mostly metaphorically transparent and literally plausible idioms. Recent studies on the other hand suggest the predominant role of the left hemisphere in idiom processing. However, these studies made use of non-transparent, unambiguous idioms. Papagno et al. (2004) showed idiom processing deficits in sentence picture matching in ten left-brain damaged right-handed subjects whereas Oliveri et al. (2004) created temporary disruption of left and right hemisphere processing using transcranial magnetic stimulation (rTMS). The results primarily showed left temporal lobe involvement, but left

Table 1
Experimental conditions with literal and idiomatic English translations

<table>
<thead>
<tr>
<th>Condition</th>
<th>Dutch Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>aL</td>
<td>Tijdens het interview raakte de minister uit de plooi.</td>
</tr>
<tr>
<td></td>
<td>During the interview became the minister/out of the pleat.</td>
</tr>
<tr>
<td>aI</td>
<td>Tijdens het wassen raakte de rok uit de plooi.</td>
</tr>
<tr>
<td></td>
<td>During the laundry became the skirt out of the pleat.</td>
</tr>
<tr>
<td>uL</td>
<td>Door de treinstaking/ zat een groep reizigers in de rats.</td>
</tr>
<tr>
<td></td>
<td>Due to the strike/sat a group of travelers in the ‘funk’.</td>
</tr>
<tr>
<td>uI</td>
<td>Door de overboeking vertrok een groep toeristen uit het hotel.</td>
</tr>
<tr>
<td></td>
<td>Due to overbooking/ left a group of tourists out the hotel.</td>
</tr>
</tbody>
</table>

Note: aI: figurative sentence with ambiguous idiom, al.: literal sentence with a potentially ambiguous idiom, uL: figurative sentence with unambiguous idiom, ul: unambiguously literal sentence.

The idioms are typeset in bold. The lines indicate presentational phrase borders. In the ambiguous conditions, the two versions of context are printed in italics. For the figurative sentences, both a literal, word-by-word, English translation is given as well as an idiomatic English translation.
frontal rTMS induced disruption to some extent as well. Interestingly right temporal rTMS had an effect as well; it facilitated subjects’ performance for both idioms and literal sentences. The authors explain this facilitation effect as disinhibition of the left hemisphere occurring after right hemisphere rTMS, based on reports of similar effects between contralateral brain areas in other task domains.

These findings suggest both hemispheres are involved in idiom comprehension but the higher degree of metaphorical transparency and/or the presence of ambiguity give prominence to the right hemisphere. This would be in line with the results of divided visual field experiments, which suggest that the right hemisphere is involved in the comprehension of ambiguous words (Burgess and Simpson, 1988; Faust and Chiarello, 1998) as well as in metaphor processing (Anaki et al., 1998; Burgess and Chiarello, 1996). The findings of Bottini et al. (1994) are consistent with a role of the right hemisphere in metaphor processing although more recently Rapp et al. (2004) found primarily left hemisphere involvement. On the other hand, neuroimaging studies on lexical semantic ambiguity comprehension in context suggest that ambiguity processing per se can also lead to right hemisphere involvement (Zempleni et al., 2007; Stowe et al., 2005a; Rodd et al., 2005). Support for an important role for meaning selection in idiom comprehension comes from Titone et al. (2002). They tested whether patients with schizophrenia can understand the figurative meaning of idioms. The results suggested that idiom comprehension was only impaired when a literally plausible alternative was available, i.e., that meaning selection caused difficulties rather than figurativeness per se. This suggests that different parts of the neural network might support the comprehension of ambiguous versus unambiguous idioms.

Materials and methods

Materials in the fMRI experiment

As described in the Introduction and illustrated in Table 1, the test materials consisted of four experimental conditions: figurative sentence with an ambiguous idiom (aI), literal sentence with an ambiguous potential idiom (aL), figurative sentence with an unambiguous idiom (uI) and unambiguously literal sentences (uL). These conditions were created using sixty-four ambiguous and thirty two unambiguous Dutch idioms, which were selected and embedded in sentence context using a number of condition matching criteria as described in the following sections. More ambiguous idioms were necessary because they occurred in two types of sentence context; see list allocation below. Different groups of native speakers participated in the off-line pre-tests.

Idioms were considered literally plausible if a plausible literal sentence could be constructed using them (see plausibility ratings below). An idiom was considered literally implausible if it was not possible to create a literally plausible sentence containing it; this could be due to syntactic or semantic ill-formedness of the idiom or implausibility of a literal interpretation in pragmatically normal circumstances. We confirmed the literal implausibility by obtaining expert judgment from a panel of seven native Dutch linguists from the University of Groningen, who rated these idioms presented without context on a one-to-five scale (1 = very easy to insert it into a plausible literal context; 5 = impossible to insert it into a plausible literal context). The average rating indicates indeed literal implausibility for this set of idioms (average = 4.6; SD = 0.29).

We checked whether the two classes of idioms were equally familiar using two ratings on a 1 to 5 scale, completed by the same nine native speakers. In the first, raters estimated how frequently they encountered or used the idioms (1 = never encountered or used; 5 = frequently heard/read/used). The two idiom groups were rated approximately equally; ambiguous idioms: 3.42 on average (SD = 1.34), unambiguous idioms 3.85 on average (SD = 1.24). On the second scale, raters indicated whether they knew the figurative meaning of the expression (1 = does not know the meaning; 5 = knows the meaning for sure). Again, the results for the two idiom groups were approximately equal; ambiguous idioms 4.63 on average (SD = 0.93), unambiguous idioms 4.88 (SD = 0.50). We decided to use both scales, as these two aspects of familiarity are not completely identical, although certainly not completely independent.

Finally, we tested the idioms to see how transparent the metaphor remains in the two groups. Twelve linguistic students, naïve to the literality of the idioms, rated this feature using a one-to-seven scale (1 = opaque idiom; 7 = transparent idiom). The two idiom groups were rated approximately equally; ambiguous idioms: 4.80 on average (SD = 0.84) unambiguous idioms 4.55 on average (SD = 0.89). These ratings suggest that the two idiom groups did not differ in this respect and that both idiom types were perceived as relatively transparent by the majority of the raters.

Having selected an appropriate set of idioms, the experimental sentences were constructed. Literal and figurative contexts were created for the ambiguous idioms whereas unambiguous idioms were presented only in an idiomatic context. The biasing context primarily preceded the idiom. Unambiguously literal sentences served as the fourth condition. Care was taken that the syntactic structures of the idioms and the sentence context surrounding the idioms used in the four conditions were matched as closely as possible. Thus the onset and the offset of the idioms within the sentences were matched across the conditions. Examples are presented in Table 1. The experimental conditions were also matched for sentence length and word frequency, which was calculated using the CELEX data base (Baayen et al., 1993). See Table 2.

Plausibility of the sentences was rated by seventy-eight native Dutch speakers (students from the University of Groningen) using a questionnaire with a one-to-five scale (1 = totally implausible, 5 = totally plausible). See Table 2. Plausibility rating lists were created in such a manner that raters encountered either the figurative or the literal meaning of the ambiguous idioms. Plausibility questionnaires were balanced with implausible filler sentences. As can be seen from Table 2, conditions did not differ in

<table>
<thead>
<tr>
<th>Condition</th>
<th>Number of sentences</th>
<th>Sentence plausibility in letters</th>
<th>Length Average word frequency</th>
<th>Sentence imageability</th>
</tr>
</thead>
<tbody>
<tr>
<td>aI</td>
<td>64</td>
<td>3.77 (0.61)</td>
<td>47 (8)</td>
<td>2.90 (0.22)</td>
</tr>
<tr>
<td>aL</td>
<td>64</td>
<td>3.90 (0.53)</td>
<td>46 (8)</td>
<td>2.87 (0.24)</td>
</tr>
<tr>
<td>uI</td>
<td>32</td>
<td>3.85 (0.65)</td>
<td>46 (8)</td>
<td>2.90 (0.23)</td>
</tr>
<tr>
<td>uL</td>
<td>32</td>
<td>4.13 (0.50)</td>
<td>47 (8)</td>
<td>2.87 (0.25)</td>
</tr>
</tbody>
</table>

Note: aI: figurative sentence with ambiguous idiom, aL: literal sentence with a potentially ambiguous idiom. uI: figurative sentence with unambiguous idiom. uL: unambiguously literal sentence.
terms of average sentence plausibility. The plausibility ratings also confirmed that the ambiguous idioms indeed have plausible literal interpretations.

Another aspect on which sentences containing idioms may vary is their imageability. In order to determine to what extent the experimental conditions differ in this respect, a rating was carried out by naive native Dutch speakers using a questionnaire with a one-to-five scale (1 = not imageable, abstract; 5 = imageable, concrete). The ratings indicated that idiomatic sentences, irrespective of the idioms’ ambiguity, were perceived as less imageable, therefore more abstract, than literal sentences. The results however suggest that the imageability of the two idiom groups was comparable. Similarly, the high imageability ratings of the ambiguous literal condition suggest that the literal context successfully disambiguated the ambiguous idioms. See Table 2. These ratings are in line with the findings of Cacciari and Glucksberg (1995), who used a mental-image production task to determine whether idioms’ concrete-literal or abstract-figurative meanings elicit more mental images. The results indicated that the majority of images obtained reflected the literal meanings.

Once the target items were constructed, as described above, thirty-two filler sentence and target word pairs were added to the materials, to be used in a relatedness decision judgment task (see under Apparatus and procedure). Half of the filler–target pairs were semantically related and half unrelated as illustrated in Table 3. Several fillers also consisted of idiomatic expressions not used in the target sentences. The semantic relationship between the sentences and words was intentionally easy to decide, so poor performance on this task would indicate lack of cooperation or attention. This task was also intended to distract subjects from the idiom comprehension task.

In order to avoid repetition effects during presentation, two experimental lists were created, which contained either the idiomatic or the literal version of an ambiguous idiom, so participants encountered only one of these versions. Hence both experimental lists contained thirty-two sentences in each condition. During allocation of the sentences to the two lists, care was taken that the aI and aL conditions in the two lists did not differ significantly in any of the sentence characteristics. Sentence order on the two lists was pseudo-randomized, so that each condition was spread evenly across the experiment with no more than two items of the same condition presented consecutively.

### Apparatus and procedure

The sentences were presented visually, phrase-by-phrase, i.e., a group of words at the same time (see Table 1), with each phrase centered on the computer screen. For programming and presentation we used E-Prime (Psychology Software Tools Inc., 2001). A projector transmitted the stimuli from the computer to a screen which was visible to the subjects via a coil-mounted mirror. The text was printed in black on a white background, with a font size which allowed for comfortable reading. Subjects were instructed to read and comprehend the sentences silently, and to carry out a relatedness decision whenever a word printed in red capital letters appeared after a sentence. Relatedness decision meant that subjects had to decide whether the words and the sentences preceding them were related in meaning. Responses were made by pressing buttons on a response box with the index and middle fingers of the right hand. This task was imposed on the filler sentences in order to check for attention and general level of comprehension of participants. We chose not to use an overt decision on the target sentences in the scanner, as the accuracy of decision and the motor responses might differ between conditions and might create a BOLD difference not related to the comprehension of the test materials per se. Furthermore, motor responses might wash out more subtle cognitively induced differences, particularly in the frontal lobes. In order to be sure that the activations seen in the fMRI results were due to understanding the sentences rather than to trying unsuccessfully to understand them, comprehension of idioms was tested right after the scanning session.

Each item was presented as follows. First a fixation cross was presented in the center of the screen, simultaneously with the start of a volume acquisition. This remained until the first phrase of the sentence appeared. Sentence onsets were jittered between 0 and 3 s which was the volume acquisition (TR) time. Each phrase remained centered on the screen for approximately 100 ms per character and was then replaced by the following phrase. After the final phrase, a fixation cross was displayed until the beginning of the following item, which was initiated by a trigger from the MR scanner. The average presentation time for the sentences was approximately 5300 ms (SD = 900 ms) for each condition\(^1\) so the fixation between the sentences (summing over pre- and post-sentence fixation time) was 15.7 s on average. Only the sentence duration periods were considered as events and modeled during data analysis (see below in Data analysis and model specification). Therefore, the total time for each fixation–sentence–fixation combination was 21 s, the time to acquire seven brain volumes. The long inter stimulus interval was chosen because the stimuli are fairly long and the necessary amount of jittering was unknown; therefore we chose to allow the hemodynamic response function (HRF) to return to baseline for each stimulus. In order to avoid effects of fatigue and lack of concentration, the experiment was divided into four runs. The order of the four runs was counterbalanced over subjects resulting in 2 lists × 2 presentation orders.

The measurements were recorded on a Philips Intera 3 T MRI at the Neuroimaging Center, University of Groningen. 336 echo planar images (T2* weighted) were acquired for each of the four runs (TR = 3000 ms, TE = 35 ms, flip angle = 90°). Each volume consisted of 46 slices covering the whole brain (slice thickness = 3.5 mm, slice gap = 0 mm, field of view 224 × 224 × 161 mm, in plane matrix size = 64 × 64).

The experiment was approved by the Medical Ethical Committee of the University Medical Center Groningen. Participants gave written informed consent prior to the experiment in accordance with the Helsinki Declaration. Before inclusion, the participants were screened for MRI incompatibility, and during scanning standard MRI safety regulations were followed.

---

1. Spaces and punctuation were also counted when calculating presentation time.

---

### Table 3

| Filler and related target | De suppoost bewaakte het schilderij. | Museum |
| Filler and unrelated target | Het hert springt over een gevallen boom. | Telephone |
Measuring participants’ behavioral performance on idiom comprehension

Participants’ performance on idiom comprehension was measured after the scanning session usually on the same day using a relatedness decision paradigm. The rationale of this task was to check whether subjects correctly chose the contextually congruent meaning of the ambiguous idioms. For this purpose, thirty literally plausible idioms were selected from the in-scanner materials and two new contexts were created for each, one biasing towards the idiomatic interpretation, the other towards the literal. Examples of the test sentences are illustrated in Table 4. Thirty sentence pairs (idiomatic versus literal context) were created; for each pair three target words were selected. The first was related to the figurative meaning, the second to the literal interpretation and the third target word was unrelated to either of the meanings. See Table 4. This produced six conditions: the idiomatic sentence followed by the congruent idiomatically related target (I-con); the idiomatic sentence followed by the incongruent literally related target (I-inc); and the idiomatic sentence followed by the unrelated target (I-unr). For the literally biasing sentences, parallel conditions were created, i.e., L-con, L-inc and L-unr.

The experimental conditions were matched for sentence structure and plausibility; plausibility ratings were carried out the same way as those described above. The average sentence plausibility for the Idiomatic sentences was 4.27 (SD = 0.49) and for the Literal sentences 4.38 (SD = 0.58). The sentence pairs usually differed only in one word. Those words were matched for length (both Idiomatic and Literal = 7 letters on average) and lemma frequency (Idiomatic = 1.32 (SD = 0.74) and Literal = 1.01 (SD = 0.85)) respectively on average; frequency data from CELEX databases (Baayen et al., 1993). The test materials were assigned to six experimental lists, containing five items in each condition. Sixty filler sentence–target word pairs were added to distract from the ambiguous items and to ensure that there was an equal number of related and unrelated sentence–word pairs in total.

The target words following the sentences belonging to the I-con/L-inc, I-inc/L-con and I-unr/L-unr conditions were also matched. Thirty words were selected for each condition. The mean lengths of the words in the different conditions, given in the order described above, were 6.8 (SD = 2); 6.0 (SD = 2); 6.6 (SD = 2); the mean lemma frequencies were 1.2 (SD = 0.7); 1.3 (SD = 0.9); 1.3 (SD = 0.7). The majority of target words were nouns in each condition. The target words’ concrete versus abstract nature was pre-tested by native Dutch speakers (approximately 10–12 students from the University of Groningen rated each word), because in the literature concreteness has been suggested to influence the hemispheric processing of word meanings. The ratings were carried out using a questionnaire with a one-to-five scale (5 = concrete; 1 = abstract). As expected, the target words were more abstract in the I-con/L-inc condition (average = 2.16; SD = 0.7) than in the I-inc/L-con (average = 3.37; SD = 1.0) and in the I-unr/L-unr (average = 3.59; SD = 1.4) conditions.

The task was implemented using a PC and subjects had to answer whether the sentence-target pairs were related or not by pressing keys on the computer with the right hand. The procedure was otherwise identical to the relatedness decision task used in the fMRI experiment (see Apparatus and procedure section above). Participants were instructed to answer as precisely and quickly possible.

Subjects

Seventeen subjects were scanned; the data from fifteen subjects were analyzed. One subject had to be excluded due to movement artifacts, and another due to inappropriate idiom comprehension performance. The participants included were healthy, native Dutch speakers (8 males and 7 females; average age 30.8 years, SD = 7.7; average educational level 16.2 years, SD = 4.0, average score on the Dutch version of the Edinburgh Handedness Inventory (Van Strien, 1992) 0.99, SD = 0.05 (almost exclusive right handedness). All subjects had normal or corrected to normal vision, normal hearing, and no history of neurological or psychiatric disorder.

Data analysis and model specification

The raw fMRI data were converted into analyze format using the MRicro software package (Rorden and Brett, 2000). SPM99 (Wellcome Department of Cognitive Neurology, London, UK) was used for spatial preprocessing, such as realignment, stereotactic normalization (2 × 2 × 2 mm voxel size) and smoothing with a 10 mm Gaussian kernel (Friston, 1994). The realigned data were checked for movement artifacts manually; translation movements bigger than 5 mm, and rotation movements bigger than 3° were rejected, leading to exclusion of one subject. Data from one additional run had to be left out, because the subject fell asleep; the other runs of this subject were included.

What we were interested in is how idioms are comprehended in a natural, sentential context, so the whole sentence duration was modeled; due to the variable length of the sentences these short epoch were modeled as events (SPM99 manual). Beta estimates were calculated for each subject for each condition, aI, aL, uI, uL (first level analysis). These beta estimates were entered into a random effects second level ANOVA analysis (Henson and Penny, 2005; Penny et al., 2003; Friston et al., 1999). To show directionality of the effects identified with this analysis, equivalent one-way T tests were generated and reported in the Results section. For those areas, which showed interaction, plots were generated showing the size and direction of the effects in order to investigate

### Table 4

<table>
<thead>
<tr>
<th>Experimental conditions in the behavioral testing</th>
<th>Target words</th>
<th>Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentences with literal and idiomatic English word-by-word translation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Het plan werd in de ijskast gezet.</td>
<td>Uitstel</td>
<td>I-con</td>
</tr>
<tr>
<td>The plan was in the fridge put.</td>
<td>Koeling</td>
<td>L-inc</td>
</tr>
<tr>
<td>The plan was put on the back burner.</td>
<td>Relatie</td>
<td>L-unr</td>
</tr>
<tr>
<td>Het eten werd in de ijskast gezet.</td>
<td>Koeling</td>
<td>L-con</td>
</tr>
<tr>
<td>The food was in the fridge put.</td>
<td>Uitstel</td>
<td>L-inc</td>
</tr>
<tr>
<td>The food was put in the fridge.</td>
<td>Relatie</td>
<td>L-unr</td>
</tr>
</tbody>
</table>

---

2 Since idiomatic sentences were rated as significantly more abstract than the literal sentences, the target words which are semantically related to these conditions are expected to differ as well.

3 An attempt was made to recruit a more varied population than university students in order to provide a more suitable group of comparison for schizophrenia patients in a latter study.
which conditions contributed to the significant interaction (see Figs. 4 and 5). These plots also show the \( T \) values of the paired contrasts calculated for these regions of interest using MarsBar (Brett et al., 2002).

**Results**

**Behavioral performance**

Participants’ performance in the scanner on the relatedness judgment task (see Apparatus and procedure section) during all runs which were included in the fMRI analysis was good. This indicates that participants were attentive and cooperative during scanning. The average error rate was 0.03 (SD=0.04).

In terms of the performance on idiom comprehension the average error rate per subject over all condition was calculated. One subject, who performed more than 2 standard deviation below the mean of the entire subject group (0.40 error rate) was excluded from the fMRI as being poor on idiom comprehension. The remaining fifteen subjects performed well on the behavioral test (average error rate across all conditions =0.17, SD = 0.06, minimum=0.03, maximum=0.25).

**fMRI results**

The areas, which showed activation at uncorrected \( p \leq 0.001 \) \((T \geq 3.30)\) and consisted of at least 10 voxels \((kE \geq 10)\) are listed in Table 5, and illustrated in Figs. 1–3. Table 5 presents the anatomical regions, showing significant changes in blood flow, with the stereotactic coordinates according to Talairach and Tournoux \((x, y, z; 1998)\), the cytoarchitectural designation according to Brodmann \((BA)\), the maximum \( T \) value \((T)\) of the contrasts and the size of each area expressed as number of voxels \((kE)\). In order to emphasize laterality effects, the hemispheric lateralization \((Lat)\) is presented in a separate column.

The figurative versus literal sentence contrast elicited activation in the left inferior frontal gyrus \((BA 47 \text{ extending to } BA 45)\), the left middle temporal gyrus \((BA 21)\) and the homologues of these areas in the right hemisphere \((BA 47 \text{ and } 21)\), though the activation in the right middle temporal gyrus \((BA 21)\) has to be interpreted in light of a significant interaction with ambiguity (see below). The literal versus figurative contrast, on the other hand, elicited left cerebellum, left occipital lobe \((cuneus \text{ and lingual gyrus})\) and right inferior frontal gyrus \((BA 13)\) activations. The ambiguous sentences versus unambiguous sentences did not elicit significantly increased activation in any brain areas even at this relatively permissive threshold. The unambiguous sentences versus ambiguous sentences lead to activation in the right insula \((BA 13)\), left basal ganglia, left cerebellum and the left precentral gyrus \((BA 6)\). The voxels showing interaction are located in the left basal ganglia and in the right middle temporal gyrus \((BA 21)\).

In order to see, whether the areas from the ambiguous/unambiguous and figurative/literal \( T \) tests overlap with the areas showing interaction, we carried out an analysis, in which the ‘main effect’ \( T \) maps were inclusively masked with the interaction \( T \) maps. This revealed that the only brain region, in which the activation has to be interpreted in light of the interaction, is the right middle temporal gyrus \((BA 21)\). Marked with bold typeface and * in Table 5.

Figs. 1–3 show the areas projected on standard anatomical T1 images \((MRljero; \ Rorden \ and \ Brett, 2000)\). The right middle temporal gyrus \((BA 21)\) is marked with a yellow circle in order to emphasize the overlap between the figurative/literal contrast and the interaction \( T \) maps.

For those areas, which showed interaction, i.e., basal ganglia and right middle temporal gyrus, plots were generated showing the

| Table 5 Anatomical regions showing significance in the main contrasts |
|---|---|---|---|---|---|
| kE | \( T \) | Lat | Anatomical region | \( x \) | \( y \) | \( z \) |
| **Frontal lobe** | | | | | | |
| 243 | 5.39 | LH | Inferior frontal gyrus | −51 | 29 | −5 47 |
| 4.86 | −53 | 25 | 2 45 |
| 19 | 3.55 | −50 | 18 | 19 45 |
| 16 | 3.72 | RH | Inferior frontal gyrus | 53 | 21 | −6 47 |
| **Temporal lobe** | | | | | | |
| 83 | 4.30 | LH | Middle temporal gyrus | −57 | −41 | −6 21 |
| 3.48 | −57 | −33 | −8 21 |
| 188 | 4.25 | RH | Middle temporal gyrus | 53 | −26 | −12 21* |
| | | | | | | |
| **Ambiguous sentences > unambiguous sentences \((aI+uI+aL+uL)\)** | | | | | | |
| Frontal lobe | 29 | 4.45 | RH | Inferior frontal gyrus/Insula | 36 | 9 | −11 13 |
| Occipital lobe | 131 | 4.65 | LH | Cuneus/Insula | −12 | −79 | 9 17 |
| | 4.09 | Lingual gyrus | −8 | −76 | 4 18 |
| | 3.48 | −14 | −70 | 2 18 |
| Cerebellum | 61 | 4.67 | L | Cerebellum | −22 | −59 | −14 |
| | 40 | 4.09 | −26 | −44 | −21 |
| | 3.91 | −28 | −36 | −27 |

**Note.** All areas reported were significant at \( p \) uncorrected \( \leq 0.001 \) \((T \geq 3.30)\), the spatial extent of activation \((kE)\) was \( \geq 10 \) voxels. Areas are presented with the stereotactic coordinates according to Talairach and Tournoux \((x, y, z; 1998)\), the cytoarchitectural designation according to Brodmann \((BA)\), the maximum \( T \) value \((T)\) and the extent \((kE)\) of the activated clusters. In order to emphasize laterality effects, the hemispheric lateralization \((Lat)\) is presented in a separate column. L: left; R: right; H: hemisphere. *The area which was significant in a main effect and in the interaction as well is marked with bold typeface. According to the masking the maximum of this area was at the 57, −26, −12 Talairach and Tournoux coordinates and the extent of the overlapping cluster was 10 voxels.
size and direction of the effects in order to investigate which conditions contributed to the significant interaction (see Figs. 4 and 5). These plots also show the $T$ values of the paired contrasts calculated for these regions of interest using MarsBar (Brett et al., 2002).

Discussion

The goal of the current study was to clarify the neural substrate of idiom comprehension using fMRI, attempting to separate the components of figurative processing versus ambiguity resolution. The test materials consisted of familiar idiomatic expressions presented in sentence context. Two types of idioms were used: literally plausible, i.e., ambiguous idioms (aI) and literally implausible, i.e., unambiguous idioms (uI). We included an unambiguous literal condition (uL) without idioms and an ambiguous literal (aL) condition as well, in which the literally plausible idioms were biased by the sentence context in which they appeared towards the literal interpretation. In order to investigate figurative comprehension we contrasted figurative sentences with literal sentences. Ambiguity resolution was studied by contrasting the ambiguous versus unambiguous sentences.

The figurative versus literal sentence contrast elicited activation in the bilateral inferior frontal gyri and in the bilateral middle temporal gyri. This suggests that figurative comprehension is more effortful than literal processing in the classical language areas and in their right hemisphere homologues. The bilateral pattern of activation clearly indicates that both hemispheres are involved in idiom comprehension and is compatible with the neuropsychological findings that both damage to the right hemisphere (Kempler et al., 1999; Van Lancker and Kempler, 1987) and damage to the left hemisphere (Papagno et al., 2004) disrupt idiom processing.

Our results are also consistent with the neuroimaging results on another form of figurative processing: the comprehension of metaphors. Bottini et al. (1994) used positron emission tomography (PET) and compared metaphorical and literal sentence comprehension. The comprehension of metaphorical sentences activated the right prefrontal cortex, right middle temporal gyrus, right precuneus and the posterior cingulate. In contrast, Rapp et al. (2004) who used event-related fMRI found no evidence for right hemisphere involvement. Their stimuli consisted of sentence pairs with either a metaphorical or a literal meaning presented visually. Subjects judged whether the sentences had a positive or negative connotation. The metaphorical versus literal sentence contrast elicited activation in the left hemisphere only: in the left lateral inferior frontal (BA 45 and 47), inferior temporal (BA 20) and posterior middle/inferior temporal (BA 37) gyri. Although Bottini et al. (1994) and Rapp et al. (2004) report opposite lateralizations, the findings from these two studies, taken together, essentially replicate the specific bilateral set of activations found in our study. As noted in the introduction, one of the issues about idiom processing is the extent to which idioms are processed in a way comparable to metaphors, despite their frozenness; some models have suggested that metaphorical processing can be skipped, since the meanings are stored, while others suggest that figurativeness remains important (Nunberg et al., 1994; Glucksberg, 1993; Gibbs,
1993). The results of the current experiment are more in line with
the latter view. This may of course depend on the extent to which
the metaphor remains transparent; we did not manipulate this factor
and the metaphorical transparency ratings of our test idioms were
relatively high.

More generally, the two areas in the left hemisphere have
been reported in a number of language localization studies. BA
47 and 45 are considered to be part of Broca’s area in a wider
sense, which is one of the classical language areas. Recent
studies suggest that these regions are involved in semantic and
pragmatic processing including sensitivity to pragmatic and
semantic violations (Stowe et al., 2005b; Kuperberg et al.,
2003). The left middle and inferior temporal gyri BA 20 and 21
are also involved in some aspects of language processing
including lexical semantic retrieval (Booth et al., 2002; Pilgrim
et al., 2002; Wiggs et al., 1999). Although less consistently,
right hemisphere activations in similar regions have been
elicited by neuroimaging paradigms which manipulated lan-
guage. The right temporal lobe was activated by increased need
for lexical semantic retrieval (Pilgrim et al., 2002; Wiggs et al.,
1999) or semantic integration and was sensitive to the
plausibility of the test materials. For example Kuperberg et al.
(2000) found that right middle and superior temporal gyri are
sensitive to semantic violations. Kircher et al. (2001) found
right temporal involvement when subject completed low Cloze
probability sentences. Homologue frontal and temporal areas are
also considered to play a role in generating the N400, which is
an electrophysiological component highly sensitive to semantic
integration difficulties (e.g., Van Petten and Luka, 2006; Halgren et al., 2002). Recent neuroimaging studies also suggest the involvement of right inferior frontal gyrus when language processing demand increases (Poldrack et al., 2001; Meyer et al., 2000).

Increased activation for literal sentences relative to figurative sentences is also seen in our study. These areas are usually not implicated in language processing: left cerebellum, left occipital lobe (cuneus and lingual gyrus) and the right inferior frontal gyrus (BA 13).

Taken together, our results suggest that both figurative and literal processing recruit cognitive resources reflected in extra activation in several brain regions, but figurative processing specifically recruits the language network. Why would idiomatic sentences elicit more activation than literal sentences in areas usually involved in normal language processing? The answer depends on how we assume idioms are processed. Several idiom processing models have been proposed (for reviews see Hillert, 2004; Titone and Connine, 1999). One aspect in which models differ is whether the literal meaning is accessed at all or accessed first. According to the standard three-stage model of non-literal language comprehension, first the literal meaning of an utterance is accessed, then the meaning of the utterance is tested against the context: if the literal meaning makes no sense, an alternative is searched for (e.g., Janus and Bever, 1985). Therefore, during idiom comprehension, first all lexical entries would be accessed for each constituent word, then, if necessary, the lexical recognition system would shift to an ‘idiom mode’.
This ‘shift to an idiom mode’ can be interpreted as need for updating/reintegrating meaning within the context. This interpretation of our data is compatible with the overlap of the current results with those studies which manipulated integrative demand in other manners (see above). For example, Zempleni et al. (submitted) found that meaning update, i.e., meaning reintegration, elicited in sentences in which context was only compatible with the subordinate meaning of ambiguous words leads to a similar bilateral neural pattern. In this study, ambiguous words, e.g., ‘bank,’ were presented in the beginning of sentences and were disambiguated after at least one intervening phrase at the end of the sentences either to the most frequent, therefore most expected, dominant meaning or to a less frequent, subordinate meaning (De advocaat werd door het oudje keurig benaderd/ingeschonken; literal English translation: ‘The lawyer/egg-flip was by the old person nicely approached/poured out’). Metaphors have also been demonstrated to increase integrative demands; metaphors which are rated off-line as plausible nevertheless elicit ERP effects which suggest that they are temporarily perceived during on-line processing as implausible (Tartter et al., 2002).

In contrast, some direct access models propose that the figurative meaning is directly accessed without pursuing literal compositional analysis. One of the direct access models, the lexicalization hypothesis advanced by Swinney and Cutler (1979), proposes that idioms are stored in the mental lexicon as individual items and are retrieved as whole words. They found that when subjects made lexical decision, they were faster on words embedded in idioms than in literal expressions suggesting that the entire idiom had already been accessed. This model would predict for our paradigm that the figurative sentences will be easier to comprehend than the literal sentences especially in areas which support lexical access and sentential processing. Note that this facilitation effect has to be seen for all idioms, in this case for both the AI and the UI conditions, as opposed to the literal conditions. We indeed found areas showing a main effect in this direction (right inferior frontal gyrus, left occipital lobe (cuneus and lingual gyrus) and left cerebellum), but since these areas are not at all part of the classical language system, we are reluctant to interpret their activation as reflecting access of idioms via a route similar to that used in lexical access, as suggested by direct access models.

There is a potential explanation, however, for the involvement of some regions in this contrast. As can be seen in Table 2, the literal sentences were rated off-line as more concrete/more imageable than the figurative sentences. This difference in imageability between the conditions may explain the main effect in some of these areas, at least in the occipital lobe (cuneus and lingual gyrus). Mental visual imagery is known to activate a similar network to the substrate of visual perception, naturally including the occipital lobe. However, studies investigating mental visual imagery have also reported right inferior frontal gyrus, though an area more lateral or closer to the insular region than in our study, and left cerebellar activation (Ganis et al., 2004).

A third model of figurative processing suggests that both figurative and literal meanings are obligatorily processed (Giora and Fein, 1999; Titone and Connine, 1999). This model easily explains the increased activation for the ambiguously idiomatic sentences (AI) relative to the unambiguous literal condition (UL), seen in the main effect of figurativeness. However, this would also predict that the ambiguous literal (AL) condition should also show additional activation relative to the unambiguous literal (UL) condition, because both meanings have to be processed in this condition as well. Therefore, both ambiguous conditions should be activated. It is not entirely clear what these models predict for the unambiguous idioms, but would presumably predict either a main effect of ambiguity or an interaction between ambiguity and figurativeness. None of the areas showed either of these patterns (see below). However, this argument relies on negative results, because the current paradigm did not primarily address this issue, so further investigation is warranted.

Moving to the contrasts between ambiguous and unambiguous sentences, ambiguous sentences did not elicit significantly increased activation in any brain areas at this relatively permissive threshold (p uncorrected ≤ 0.001; T ≥ 3.30; kE ≥ 10 voxels). It is remarkable that the unambiguous sentences elicited significant activation rather than the ambiguous sentences. The areas involved were the right insula (BA 13), left basal ganglia, left cerebellum and the left precentral gyrus (BA 6). It is difficult to explain why the unambiguous conditions would elicit more activation than the ambiguous conditions as there is no model which would propose an additional cognitive component in the processing of these sentences. However, several studies have indicated that both figurative and literal meanings are obligatorily activated during idiom comprehension (Giora and Fein, 1999; Titone and Connine, 1999). Therefore, one might pursue the idea that during comprehension of ambiguous conditions one of the alternative meanings has to be suppressed (Galinsky and Glucksberg, 2000; Gernsbacher and Robertson, 1999). However, this would only explain the pattern we see, if neural inhibition is manifested as task relevant deactivation, a proposal which is currently under considerable debate (Shmuel et al., 2006; Stefanovic et al., 2004; M.-Z. Zempleni et al. / NeuroImage 34 (2007) 1280–1291 1289
Hamzei et al., 2002; Raichle et al., 2001; Gusnard and Raichle, 2001; Binder et al., 1999; Shulman et al., 1997; Kawashima et al., 1995). Although inhibition has also been claimed to elicit deactivation, inhibitory processes have more frequently been found to evoke positive BOLD response. In particular the insula was found to be activated in tasks requiring neural suppression (e.g., suppressing thoughts; Wyland et al., 2003).

Lastly we consider the areas which showed significant interaction between figurativeness and ambiguity. The majority of the voxels showing interaction are located in the left basal ganglia and in the right middle temporal gyrus. The basal ganglia were significant in the (al–aL)<(ul–uL) contrast. We do not have an interpretation for the meaning of this contrast, especially in an area usually not implicated in language processing; we report it in the results section in order to provide the full data (see Fig. 4). We will not discuss it further though. The right middle temporal gyrus showed an interaction in the (al–aL)> (ul–uL) contrast. This is interesting, because this area also showed a main effect of figurativeness, meaning that the activation has to be interpreted in light of this interaction. In this case, what we see is a very clear difference between the two ambiguous conditions, with the figurative one being more activated than the literal one (T = 5.58 in the al>aL contrast), whereas the two unambiguous conditions were not different (see Fig. 5). This suggests that the right temporal lobe is more sensitive to ambiguous idioms than to unambiguous idioms. These results are consistent with earlier findings that damage or TMS knock out may selectively impair ambiguous idiom comprehension in the right hemisphere (Kempler et al., 1999; Van Lancker and Kempler, 1987) and unambiguous idiom comprehension in the left hemisphere (Oliveri et al., 2004; Papagno et al., 2004). However, this conclusion would be definitely stronger if we had seen an interaction in the left temporal lobe as well, in which selective sensitivity for the unambiguous idioms would be seen.

Conclusions

The major finding of the current study is that figurative comprehension involving idioms is supported by bilateral inferior frontal gyri and bilateral middle temporal gyrus, similar to the neural substrate of metaphor comprehension reported by earlier studies. The pattern of activation is consistent with the three-stage model of metaphor comprehension, which proposes that the literal meaning is always computed and when it does not fit into the context an alternative figurative meaning is sought. Additionally, we found evidence that the literal plausibility of idioms, i.e., ambiguity, affects the neural network involved; the right temporal lobe seems to be particularly involved in the processing of ambiguous idioms.

Acknowledgments

We would like to thank the CLCG research group from the University of Groningen and the ‘Activation Meeting’ discussion group from the Neuroimaging Center, University of Groningen for commenting on the experimental design and the results. Special thanks to Ingeborg Prinsen, Elzerieke Hilbrandie and Charlotte Wunderink for helping to create the test sentences. Remco Renken is a Hazewinkel-Beringer fellow; Monika-Zita Zempleni was supported by an Ubbo Emmius scholarship provided by the Faculty of Arts, University of Groningen.

References


