On the Tip-of-the-Tongue: Neural Correlates of Increased Word-finding Failures in Normal Aging

Meredith A. Shafto[1,2], **Deborah M. Burke**[2], **Emmanuel A. Stamatakis**[1], **Phyllis P. Tam**[1,2], and **Lorraine K. Tyler**[1]

1 *University of Cambridge, UK*

2 *Pomona College, Claremont, CA*

Abstract

Tip-of-the-tongue (TOT) experiences are frustrating word-finding failures where people are temporarily unable to produce a word they are certain they know. TOT frequency increases with normal aging during adulthood, and behavioral evidence suggests that the underlying deficit is in retrieving the complete phonology of the target word during production. The present study investigated the neural correlates of this phonological retrieval deficit. We obtained 3-D T1-weighted structural magnetic resonance images (MRI) for healthy participants between 19 and 88 years old and used voxel-based morphometry to measure gray matter density throughout the brain. In a separate session, participants named celebrities cued by pictures and descriptions, indicating when they had a TOT, and also completed Raven's Progressive Matrices (RPM), a task that does not involve phonological production. The number of TOTs increased with age and also with gray matter atrophy in the left insula, an area implicated in phonological production. The relation between TOTs and left insula atrophy cannot be attributed to the correlation of each variable with age because TOTs were related to insula atrophy even with age effects removed. Moreover, errors on the RPM increased with age, but performance did not correlate with gray matter density in the insula. These results provide, for the first time, an association between a region in the neural language system and the rise in age-related word-finding failures and suggest that age-related atrophy in neural regions important for phonological production may contribute to age-related word production failures.

INTRODUCTION

The effect of normal aging on language is characterized by declining performance on some language functions and well-maintained performance on others. Nowhere is this pattern more evident than in the contrast between retrieval of the semantic components of words and retrieval of phonological components. Retrieval of word meaning as measured, for example, on tests of vocabulary, increases during adulthood with little decline until age 75 years or older (Burke & Shafto, 2008). On the other hand, retrieval of the phonological form of a word declines during adulthood. The tip-of-the-tongue (TOT) experience is a common and dramatic word-finding failure where a person is temporarily unable to produce a well-known word. TOTs increase in frequency during adulthood (Burke & Shafto, 2004) and become one of older adults' most irksome and distressing cognitive problems (Lovelace & Twohig, 1990).

Locus of Tip-of-the-Tongue Experiences: Phonological Retrieval Deficits

TOTs are a valuable source of information about the organization of the language system because they represent a phonological retrieval deficit when semantic information about the

word has been accessed. They are consistent with models of language production that postulate a hierarchical organization of semantic, lexical, and phonological representations corresponding to a word: Language production begins with semantic activation followed by selection of a lexical representation and retrieval of its phonological components (Levelt, 2001; Rapp & Goldrick, 2000; MacKay, 1987; Dell, 1986). TOTs occur when semantic and lexical information have been selected, producing a strong feeling of knowing, but phonological retrieval is insufficient for computation of the complete phonological code (Burke, MacKay, Worthley, & Wade, 1991).

William James described the TOT state as "a gap that is intensely active. A sort of wraith of the word is in it" (James, 1890, p. 251). The "wraith" in part consists of *partial phonological information* that often is available to a person in the TOT state such as the initial letter or number of syllables (Burke et al., 1991; Brown & McNeill, 1966). The complete phonology, however, remains inaccessible until the TOT is resolved either spontaneously by the word popping into mind or by consulting an external source. Although older adults experience more TOTs than young adults both in everyday life and in the laboratory, they report less partial phonological information. This is consistent with the view that TOTs are caused by a phonological retrieval failure that worsens in old age so that even partial phonological information becomes less available (Burke et al., 1991).

A phonological locus for the deficit in TOTs is supported by experimental evidence. Using definitions of rare words to induce TOTs in the laboratory (e.g., *What is the term for formally renouncing the throne?*), young and older adults in a TOT state pronounced prime words (e.g., *abstract*) with sound segments that overlapped some of the sounds of the target word (e.g., *abdicate*). Pronouncing prime words increased resolution of the TOT when the definition was repeated (White & Abrams, 2002; James & Burke, 2000). Similar effects were observed when participants produced a homophone of a proper name in response to a definition (e.g., *The hard stone, as of the plum or cherry, which contains the seed is called the p____*). Prior production of the homophone reduced TOTs and improved production of a phonologically identical celebrity name (e.g., [Brad] *Pitt*), with a larger benefit for older than young adults (Burke, Locantore, Austin, & Chae, 2004).

These behavioral findings have been explained within a spreading activation model wherein connections among representations weaken with infrequent use or aging of the person, reducing the transmission of excitation to representations so that they become difficult to access. Phonological representations are more vulnerable to such transmission deficits than semantic representations because the representation of meaning is characterized by greater redundancy than the representation of sound (Burke & MacKay, 1997; Burke et al., 1991). Weak phonological connections are strengthened by production of the relevant phonology, which aids retrieval, consistent with phonological priming effects (James & Burke, 2000). There is no evidence that TOTs involve deficits at the semantic level: Semantic information is fully available during a TOT and production of semantically related prime words does not reduce TOTs (Cross & Burke, 2004; Meyer & Bock, 1992). There is also no evidence that TOTs involve deficits at the articulatory level because TOT targets can be neither spoken nor written (e.g., Rastle & Burke, 1996; Meyer & Bock, 1992). The involvement of both production modalities differentiates TOTs from speech production failures caused by deficits in programming of speech motor movements as in speech apraxia where writing of words is preserved (Dogil et al., 2002). Unlike people in a TOT state, people with apraxia of speech know the sounds of the word, but they are unable to compute the program to articulate them (Hillis et al., 2004).

The Neural Underpinnings of Age-related Word-finding Failures

Both cross-sectional (Sowell et al., 2003; Good et al., 2001) and longitudinal studies (Resnick, Pham, Kraut, Zonderman, & Davatzikos, 2003) demonstrate that age-related gray matter atrophy is widespread, although neural regions differ in the degree of atrophy. Previous research has linked regional atrophy with specific age-related cognitive changes (e.g., Raz, Gunning-Dixon, Head, Dupuis, & Acker, 1998), but this research has largely examined executive function and working memory, and has not addressed the link between specific language deficits and age-related structural changes. Thus, despite considerable research on the cognitive mechanisms underlying TOTs, little is known about their neural substrate in young or older adults. The three extant functional magnetic resonance imaging (fMRI) studies of TOTs tested young adults (Maril, Simons, Weaver, & Schacter, 2005; Kikyo, Ohki, & Sekihara, 2001; Maril, Wagner, & Schacter, 2001), and focused on the role of cognitive control mechanisms in identifying and resolving TOTs rather than the language-specific mechanisms that cause TOTs. Maril, Schacter, and colleagues found that TOTs selectively increased activation of the anterior cingulate cortex (ACC), right middle frontal, and right inferior frontal cortex compared to correct responses or don't know responses. They argued that this is consistent with an ACC–prefrontal cortex (PFC) cognitive control circuit, which is engaged in monitoring conflict. At TOT onset, the strong feeling of knowing the word conflicts with the inability to retrieve it, and during the TOT, available partial information may conflict with attempts to retrieve the target if the partial information is incorrect or involves related words (Maril et al., 2001). Selective activation of this ACC–PFC network occurs during TOTs, but not during retrieval attempts for words that lack a strong feeling of imminent recall (Maril et al., 2005).

Although the cognitive control processes associated with an ACC–PFC network seem important in coping with the TOT state, a change in the functioning of this network is unlikely to be the driving factor which causes TOTs. The behavioral evidence suggests that nonoptimal functioning of a specific component of language processing—retrieval of phonology—is the main cause of TOTs and predicts that the neural underpinnings of TOTs should involve regions responsible for phonological retrieval. A recent meta-analysis of 82 neuroimaging studies of word production identified neural regions which were reliably associated with different components of language production (Indefrey & Levelt, 2004). An association between a neural region and a specific language processing component was obtained when the region was active during experimental tasks requiring this component of production but was not active in tasks not requiring the component. The authors identified two regions associated with phonological retrieval: the left posterior superior and middle temporal gyri (Wernicke's area) and the left anterior insula. Previous studies have emphasized the role of Wernicke's area in the comprehension rather than production of speech, although patients with damage to this region sometimes produce "neologisms," combinations of speech sounds that are not real words, suggesting that it may also play a role in phonological output. Lesion mapping studies have confirmed that Wernicke's area is not reliably associated with phonological impairments (Bates et al., 2003), unlike lesions in the insula which consistently co-occur with phonological output deficits.

However, the exact functions of the insula in word production are controversial. Insula damage has been linked to phonological retrieval deficits in patients with Alzheimer's disease (Harasty et al., 2001) and stroke (Cereda, Ghika, Maeder, & Bogousslavsky, 2002), but the insula is most commonly linked to deficits at the motor planning level. In an influential lesion-overlap study, Dronkers (1996) reported that 25 brain-damaged patients with speech apraxia all had lesions in the left anterior insula, but patients without speech apraxia did not have lesions in this region. Dronkers concluded that the left anterior insula is critically involved in articulatory planning, a conclusion supported by results from a positron emission tomography (PET) study

with healthy participants (Wise, Greene, Büchel, & Scott, 1999). Hillis et al. (2004) recently challenged this conclusion in a study of stroke patients with or without insular damage. They found no association between insula damage and apraxia of speech. They concluded that apraxia of speech and damage to the insula are both caused by large middle cerebral artery strokes, but that they are not causally related to each other.

In sum, previous behavioral research indicates that TOTs are caused by a phonological retrieval deficit that worsens with age, and the present study tests for the neural underpinnings of this deficit. Based on previous findings, we hypothesize that age-related atrophy in regions important for phonological retrieval will be affiliated with age-related increases in TOT states. Previous studies suggest the left insula as a primary candidate region. An association between insula atrophy and TOTs would implicate phonological rather than articulatory deficits because the failure in TOTs is at a level prior to articulation. Finally, if the association between TOT performance and atrophy in regions critical for phonological retrieval reflects phonological deficits rather than general cognitive decline, atrophy in these regions should not be related to performance on a task which does not require phonological production. Thus, we also test participants on a nonverbal task sensitive to aging, Raven's Progressive Matrices (RPM; Raven, 1958).

To test these hypotheses, we correlated gray matter density across whole-brain gray matter images with two tasks known to show age-related performance deficits: a TOT task, which involved naming famous faces, and the RPM, which provides a nonverbal assessment of fluid intelligence. The face naming task was chosen because, although TOTs occur for a wide range of word types, proper names reliably elicit more TOTs than other words and show a larger age difference in TOT frequency (Evrard, 2002; Rastle & Burke, 1996; Burke et al., 1991). Phonological priming of proper names reduces TOTs more for older than young adults, suggesting that phonological retrieval failures contribute to the age-related deficit (Burke et al., 2004). Thus, we expected that inducing TOTs for proper names would provide the best measure of age-related variability in TOT rates.

To examine the relationship between TOT performance and region-specific atrophy, we employed voxel-based morphometry (VBM), which is an automated technique allowing identification of regional gray matter differences on a voxel-by-voxel basis. We have used a similar voxel-based technique in recent studies (Tyler, Marslen-Wilson, & Stamatakis, 2005a, 2005b), in which we have correlated whole-brain signal intensity values for a wide variety of brain-damaged patients with their scores on behavioral tasks, and have revealed the remarkable sensitivity of the method to structure–function relations. This method offers important advantages over typical neuropsychological studies examining the link between brain damage and behavioral impairment. These studies typically employ a factorial design, categorizing patients on the basis of the presence or absence of damage in a specific brain region, and the presence or absence of a behavioral deficit. The method we employ increases the sensitivity to changes in tissue integrity and behavioral performance by avoiding binary categorization and employing a correlational design. In the present study, we correlate continuous data on neural integrity in the form of gray matter probability scores at each voxel, with behavioral performance in the form of proportional scores (e.g., proportion of TOTs) for each participant. This allows us to examine the relationship between developmental brain changes across the adult lifespan and gradually increasing TOT rates.

A second disadvantage of previous methods for evaluating structure–function relations is that they often limit their analysis to specific brain areas. Our voxel-based approach avoids making any a priori assumptions about intact or damaged tissue, thus increasing the range of values included in the analysis and the power of the inferences we can draw. We applied the technique by correlating performance on the behavioral tasks with gray matter density at each voxel in

a segmented gray matter image. We used this whole-brain analysis to motivate a region-of-interest (ROI) analysis in an area important for phonological retrieval, namely, the left insula. We predicted that TOT scores, but not RPM scores, would correlate in the left insula, and that both behavioral scores would correlate with age.

METHODS

Participants

Participants were 46 healthy adults 19 to 88 years old ($M = 55.5$, $SD = 19.97$) recruited from the University of Cambridge and surrounding community. All participants had normal or corrected-to-normal hearing and vision during behavioral testing. For structural image acquisition in the MRI scanner, no participants had audiometer results indicating hearing loss that would impair their ability to follow the directions of radiographers while inside the scanner. All participants scored 26 or above on the Mini-Mental State Exam (MMSE; Folstein, Folstein, & McHugh, 1975; $M = 29.3$, $SD = 1.13$), and scores did not correlate significantly with age. Major exclusion criteria for participation included MR contraindications, neurological or hormonal disorders, recent treatment (within 1 year) for psychiatric disorders, major head trauma, stroke, dyslexia (self-report), bilingualism (self-report), and left-handedness (self-report or less than 50% right-handed on the Edinburgh Handedness Inventory; Oldfield, 1971).

Education was assessed by assigning self-reported highest educational qualification to one of five ranked categories in the British educational system (in order from lowest to highest qualification): (1) any education up to school certificate/O-levels, with no vocational training; (2) completed A-levels up to some university, with no vocational training; (3) any education below first degree including vocational training; (4) completed first degree; (5) completed any formal education after receiving first degree. A one-way analysis of variance revealed no age difference among the five education levels ($F < 1$), and there was no correlation between age and maximum education level.

Behavioral Tasks

Participants completed behavioral sessions totalling 2 to 2.5 hours, which involved screening and background tests, as well as the TOT task. Tasks were divided across two or three testing sessions in order to reduce the risk of fatigue. These sessions were conducted individually and involved the administration of the TOT task and a number of background measures, including a questionnaire to assess current and past health conditions and other potential exclusion criteria, the Edinburgh Handedness Inventory (Oldfield, 1971), and an audiometric exam to assess hearing. Additionally, participants were given several tasks to assess cognitive and language abilities. This included the MMSE (Folstein et al., 1975), which is a screening test for dementia, and the National Adult Reading Test (NART; Nelson, 1982), which provides a measure of word knowledge and involves pronouncing 50 irregularly spelled words (e.g., *cellist*, *syncope*). Participants also completed measures of digit span forward and backward (Wechsler, 1987), a standardized 40-item vocabulary test (Shipley, 1946), and the Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983), which involves naming 60 simple line drawings. Finally, participants completed the RPM (Raven, 1958), a 60-item abstract reasoning task, which provides a nonverbal measure of fluid intelligence. In sum, these background measures provided data on participants' cognitive abilities, including working memory, vocabulary, picture naming, and fluid intelligence.

The TOT task took approximately 20 to 30 min to complete, and was administered as part of the behavioral sessions, using a method similar to that used in previous TOT studies (Burke et al., 1991). Stimuli consisted of 68 pictures of public figures from different eras (18th century

to present) and different categories (e.g., movie stars, politicians, sports figures, authors, etc.), each with a short written description (see Figure 1 for an example). These stimuli were selected from a larger set on the basis of familiarity and ability to induce TOTs in pilot testing with 10 young and older adults. Stimuli were presented in the same randomized order one at a time on a computer screen using Microsoft PowerPoint. For each trial, participants responded with the name of the person (Know response), said they did not know the name (Don't Know response), or said they were having a TOT for the name (TOT responses). Responses were not timed, but if participants were unresponsive for several seconds, the experimenter prompted them for an answer choice. As in previous studies (e.g., Burke et al., 1991), a TOT was scored when the participant indicated they were having a TOT, even if it was resolved before the next trial. For each TOT response, provided it was not yet resolved, participants were asked if they had any partial information about the name they were trying to retrieve (first letter or sound of the surname) or if any alternative names were coming to mind. Following this, the experimenter presented the correct name and asked participants to confirm that the name they were having a TOT for was the correct name. Only TOTs for correct names were counted as TOTs in the analysis. The experimenter recorded their responses including any partial information or alternative names.

Structural Imaging

In addition to behavioral sessions, all subjects were scanned on a 3-T Bruker scanner (with body gradients) using a T1-weighted SPGR sequence with the following parameters: TR = 19.18 msec, TE = 5.0 msec and T1 echo time = 5.0 msec; pulse angle 25°; matrix size $256 \times 220 \times 180$ mm; voxel size $0.859 \times 1 \times 0.703$ mm^3 field of view $25.6 \times 22 \times 18$ cm.

Imaging Analysis: Preprocessing

The T1 images were preprocessed using SPM2 (Wellcome Department of Cognitive Neurology, London, UK) utilizing the optimized VBM protocol (Good et al., 2001). This method was developed to improve segmentation of images into the gray matter, white matter, and CSF. Initially, the images were spatially normalized and segmented into probabilistic gray matter, white matter, and CSF. Template images specific to our data were constructed from the normalized segmented images and our raw images were normalized again and segmented using these data-specific templates. The final preprocessing step involved smoothing using a 12-mm isotropic Gaussian kernel, in order to make the data conform more closely to the Gaussian field model underlying the statistical procedures used for making inferences, and to render the data more normally distributed (by the central limit theorem) (Good et al., 2001). These images were then analyzed using the General Linear Model in SPM2. Multiple linear regression was used in each analysis to identify clusters with a significant correlation between gray matter density and behavioral scores (e.g., proportion of TOTs) across participants. We thresholded the statistical parametric maps at $p < .001$ (unless otherwise specified), uncorrected at the voxel level, and report maxima for clusters that survive a random field correction ($p < .05$) for gray matter volume across the entire brain. Gray matter was selected as a mask using WFU Pick Atlas (Maldjian, Laurienti, Kraft, & Burdette, 2003), which is based on the Talairach and Tournoux (1988) stereotactic atlas. All analyses also included a measure of total volume (total gray matter + total white matter + total CSF) as a confounding covariate.

RESULTS

Behavioral Results

Scores on the Shipley vocabulary test (Shipley, 1946) increased with age ($r = .59$, $p < .001$), as did scores on the NART (Nelson, 1982) ($r = .58$, $p < .001$). The NART performance indicated that older adults made fewer pronunciation errors than younger adults, probably because older adults have larger vocabularies which would aid them in knowing how to pronounce the low-

frequency irregularly spelled words, which comprise the NART. Supporting this, the correlation of NART errors and age was no longer significant when Shipley scores were partialled out. Age was not correlated with digit spans forward or backward (Wechsler, 1987), or with score on the BNT (Kaplan et al., 1983).

Age was positively correlated with number of TOTs ($r = .43$, $p < .01$), negatively correlated with the proportion of "Know" responses ($r = -.53$, $p < .001$), and positively correlated with "Don't Know" responses ($r = .33$, $p < .05$). In keeping with previous research (e.g., Esposito, Kirby, Van Horn, Ellmore, & Berman, 1999), age correlated negatively with RPM scores ($r = -.35$, $p < .05$), but TOT rates and RPM scores did not correlate significantly with each other. Additionally, TOT rates did not correlate significantly with measures of vocabulary knowledge, as measured by the Shipley vocabulary test or the NART.

Imaging Results

A VBM analysis correlating age and gray matter density at each voxel in the segmented gray matter images revealed that age correlated highly with gray matter density in a wide range of regions (see Figure 2A). A measure of mean gray matter (in the entire brain) correlated negatively with age ($r = -.89$, $p < .001$; see Figure 2B), confirming the wide-ranging age-related gray matter atrophy in regions previously identified in other studies (e.g., Sowell et al., 2003). There was extensive age-related atrophy in the frontal, parietal, temporal, and occipital cortex bilaterally with relative sparing of the middle and inferior temporal cortex.

An analysis correlating TOT scores for each participant with their gray matter density across the whole brain on a voxel-by-voxel basis revealed a bilateral pattern of negative correlation including the superior temporal gyrus, rolandic operculum, and precentral gyri. The most significant correlation was in a left-lateralized cluster (5244 voxels, peak voxel, $x = -47$, $y = -2$, $z = 11$; Z-score = 4.31), which included an extensive region of the left insula (BA 48) and extended laterally to the rolandic operculum and posteriorly to the anterior tip of the superior temporal cortex. There were also smaller clusters of correlation in the right insula (BA 47), bilateral superior temporal gyrus (BA 22, 38/42), and cingulate gyrus (BA 24).

Because RPM performance and TOT rates both correlate with age, regions of correlation in gray matter, which are common to both, may reflect a general aging factor. In particular, because RPM performance does not require phonological production, regions of correlation in common with TOT rates are unlikely to be specific to phonological retrieval. Therefore, the TOT analysis was repeated with RPM error rates as a covariate, yielding regions which uniquely correlate with TOT performance (see Figure 3A). This analysis revealed two significant clusters, with the largest including the left insula (BA 48, peak voxel, $x = -47$, $y = -2$, $z = 11$; Z-score = 4.07), and also extending to include the rolandic operculum, Heschl's gyrus, and superior temporal gyrus. Figure 3B depicts a plot at the peak statistical voxel of this analysis ($x = -47$, $y = -2$, $z = 11$; Z-score = 4.07), which confirms the negative correlation between gray matter density and proportion of TOTs. Importantly, RPM scores did not significantly correlate here, demonstrating a selective correlation for a task which required phonological production. A Williams (1959) test shows that the two correlations are statistically different, $p < .05$. In addition to this cluster including the left insula, there was another smaller cluster of correlation in the cingulate gyrus (BA 24). At a more conservative threshold ($p < .0001$), only the cluster which included the insula remained significant.

A final analysis examined the correlation between gray matter density and TOT scores when age was factored out. Although we expected that age would be the primary determinate of variation in both gray matter density and TOT rates, the cognitive model we adopted predicts that phonological retrieval failures underpin TOTs for both young and older adults (e.g., James & Burke, 2000). Thus, left insula integrity may be related to TOT rates across the lifespan,

with higher TOT rates in old age reflecting a greater reduction in gray matter density due to age-related atrophy. To test for this possibility, we carried out a correlation between TOT scores and gray matter density, partialling out variance due to age. At a lowered threshold of $p < .01$, this analysis resulted in a significant cluster with a maximum in the left postcentral gyrus (BA 40/43, peak voxel $x = -53$, $y = -14$, $z = 19$, Z-score = 3.30) which, as in the previous analyses, included the left insula, rolandic operculum, and Heschl's gyrus. Additionally, as with the previous analyses, there was a smaller cluster of correlation in the cingulate gyrus (BA 24). Taken together, these analyses demonstrate that TOT rates correlate with gray matter density in the insula, and this correlation cannot be fully explained by a global aging effect.

The whole-brain analyses correlating gray matter density with TOTs each resulted in a significant cluster which included the left insula. Because of these findings and prior evidence for the role of the insula in phonological retrieval, we carried out subsequent analyses with the left insula as an ROI. On each participant's spatially normalized, probabilistic gray matter image, we defined an ROI around the left insula using the WFU Pick Atlas (Maldjian et al., 2003). Confirming the presence of relevant age-related atrophy, mean gray matter density in the left insula correlated negatively with age ($r = -.81$, $p < .001$). We then correlated gray matter density for each voxel within the ROI with TOT scores. As predicted, proportion of TOTs correlated negatively with gray matter density in the left insula. Figure 4 depicts the plot at the peak statistical voxel of this cluster ($x = -45$, $y = -2$, $z = 11$; Z-score = 4.33), which confirms the correlation between gray matter density in the left insula and proportion of TOTs. Importantly, RPM scores did not significantly correlate here, demonstrating a selective correlation for a task which required phonological production. A Williams' test showed the two correlations were statistically different ($p < .05$, one-tailed). Finally, a significant cluster of negative correlation remained in the left insula when the correlation was repeated, partialling out RPM ($x = -45$, $y = -2$, $z = 11$; Z-score = 4.07) or age ($x = -45$, $y = -3$, $z = 11$; Z-score = 3.26); voxel level $p < .01$, uncorrected, cluster survived a random field corrected $p < .05$.

To confirm that RPM scores could elicit significant correlations with relevant cortical regions, but these regions would not include the insula, we correlated RPM error scores with voxel-by-voxel gray matter density across the entire brain. This analysis revealed a number of significant clusters primarily in the right hemisphere, including the fusiform and parahippocampal gyrus, precentral and medial frontal gyri, together with a small cluster in the left middle frontal gyrus (see Figure 5A). These regions are consistent with those shown to be active during RPM performance in previous PET and fMRI studies (e.g., Kroger et al., 2002; Christoff et al., 2001; Esposito et al., 1999; Prabhakaran, Smith, Desmond, Glover, & Gabrieli, 1997). Figure 5B shows the plot at the peak statistical voxel in the whole brain ($x = -25$, $y = 0$, $z = 51$; Z-score = 4.19), which confirms the negative correlation between gray matter density and proportion of RPM errors. Importantly, TOT scores did not significantly correlate here. A Williams test showed the two correlations were statistically different ($p < .05$, one-tailed).

DISCUSSION

These results provide the first evidence for the neural correlates of the age-related increase in frequency of TOTs during adulthood, and support findings from behavioral research that phonological retrieval deficits underpin TOTs across the lifespan. TOTs increased with age-related gray matter atrophy in the left insula, but TOT frequency correlated with insula atrophy even with the effect of age removed in both the whole-brain and ROI analyses. Thus, this correlation cannot be attributed to the separate correlations of TOTs and insula atrophy with age because insula atrophy is uniquely related to the probability of TOTs independent of aging. Moreover, performance on a nonverbal test, which does not involve phonological retrieval (i.e., RPM), declined with age but did not correlate with insula atrophy. The correlation of TOTs and insula atrophy remained when RPM performance was covaried, indicating that

insula atrophy is not simply associated with more general cognitive decline involving abilities independent of language. Taken together, these analyses demonstrate the relationship between left insula atrophy and age-related increases in TOT rates.

Cognitive models locate the deficit in TOTs as a failure of phonological retrieval (Burke & Shafto, 2004; Burke et al., 1991, 2004; White & Abrams, 2002; James & Burke, 2000). PET data from healthy subjects identify the insula as a region important in phonological processing (e.g., Blank, Scott, Murphy, Warburton, & Wise, 2002), and the present results support the role of the insula in phonological retrieval processes during language production and in their breakdown, as revealed in the occurrence of TOTs. The present study does not, however, rule out the possibility that the insula plays a role in articulatory planning (Dronkers, 1996), although evidence from aging studies provides little support for such a role (see also Hillis et al., 2004). There is no evidence that articulation declines in normal aging despite previous findings that insular cortex is one of the neural regions most impacted by age-related loss of gray matter (Resnick et al., 2003; Good et al., 2001). In the present study, correct pronunciation of low-frequency, irregularly spelled words on the NART improved with aging. This age-related improvement reflected older adults' superior vocabulary knowledge because, with vocabulary test scores covaried, there was no relation between age and NART scores. These findings provide no evidence that aging is related to articulation failures. Moreover, the link between insula atrophy and increased TOTs is not mediated by articulation failures because they play no role in causing TOTs: TOTs occur when no articulation is required as when production is written (Rastle & Burke, 1996). The beneficial effect of priming phonologically related words on resolution of TOTs occurs, regardless of whether the prime word is articulated or read silently (Abrams, White, & Eitel, 2003). If TOTs involved an articulatory planning failure, then articulating a phonologically related word should benefit TOT resolution more than silently reading it.

Investigations of the role of the insula in language have concentrated on processes involved in motor planning and articulation, but it is difficult to isolate articulation from phonological processes. Measures in some studies implicating the insula can be interpreted as reflecting phonological retrieval, articulation, or both, for example, counting and reciting nursery rhymes (Blank et al., 2002), speech fluency (Bates et al., 2003), the presence of mutism (Habib et al., 1995), and picture naming and verb generation (Etard et al., 2000). The insula is characterized by a wide range of reciprocal cortical connections and functional activities beyond language production (Cereda et al., 2002), and the left insula may play a more general coordinative or selective role involving multiple output processes. Neuropsychological research suggests that, given the insula is highly interconnected with a number of other regions in both temporal and frontal lobes (Augustine, 1996), the effect of damage to the insula is likely to vary with its location within the insula (Harasty et al., 2001). The present results, however, link atrophy of an extensive area of the insula to word-finding failures. The results support the role of the left insula in phonological rather than articulatory processing, but further research is needed to determine whether phonological processes vary regionally within the insula.

Our results raise similar issues for another region, the left rolandic operculum, which was consistently included in clusters where the left insula correlated with TOT rates, even when age or RPM score was factored out. As with the insula, studies of the rolandic operculum have focused on its role in articulatory impairments, including both acquired (Tonkonogy & Goodglass, 1981) and developmental deficits (Sommer, Koch, Paulus, Weiller, & Büchel, 2002). However, recent research suggests that the rolandic operculum is involved in other aspects of language production, including syntactic encoding (Indefrey et al., 2001). The possibility that the rolandic operculum is involved in phonological retrieval has not been widely examined, but our results suggest this may be the case, given the weight of evidence that TOTs reflect phonological rather than articulatory failures.

A second region related to TOTs that survived when the effect of age or RPM scores was covaried was a small cluster in the ACC. In fMRI studies, both the PFC and the ACC have been shown to become activated during TOT states (Maril et al., 2001, 2005) or prolonged word retrieval (Kikyo et al., 2001). Maril and colleagues suggested that the ACC is engaged during a TOT as part of a cognitive control system in which the ACC interacts with the PFC in monitoring for situations requiring executive control functions that are implemented by the PFC (Cohen, Botvinick, & Carter, 2000; Gehring & Knight, 2000). When a TOT occurs, these control functions trigger processes aimed at resolving the TOT, for example, evaluating partial information that becomes available. The correlation between ACC gray matter density and TOT frequency implies that this control system is relevant to the initial cause of word retrieval failures, but Maril et al. discuss the influence of the ACC on right-lateralized regions that are responsible for strategic retrieval attempts after a TOT occurs. A different perspective on the role of the ACC is suggested by recent connectivity analyses on fMRI data (Stamatakis, Marslen-Wilson, Tyler, & Fletcher, 2005), which show that the ACC modulates the relationship between left frontal and temporal regions during language processing, and that this modulatory function varies with the task. In the study reported by Stamatakis et al., the specific task involved morphophonological segmentation. This type of modulatory control over the relationship between the neural regions within the language system may be particularly important for retrieval success during a TOT task, where target names are selected to be difficult to retrieve. Further research using a time-sensitive methodology (such as EEG or MEG) is needed to investigate the role of the ACC in lexical retrieval at onset of a difficult retrieval process and before production of a TOT or correct response.

Conclusions

The current study demonstrates for the first time a neural correlate of increased word-finding problems in old age. The frequency of TOTs was positively correlated with adult age and was negatively correlated with gray matter density in the left insula even with the effects of age removed. These results increase our understanding of the role of the left insula in language production by demonstrating that insula atrophy is linked to increased phonological retrieval deficits. The results highlight the importance of isolating different components of word production using paradigms based on cognitive models of word production and behavioral evidence. This approach is essential for differentiating phonological and articulatory processes and clarifying the functional relation of insula to each. With respect to aging, the current findings demonstrate that region-specific gray matter atrophy is associated with a specific age-related cognitive deficit.

References

Abrams L, White KK, Eitel SL. Isolating phonological components that increase tip-of-the-tongue resolution. Memory & Cognition 2003;31:1153–1162.

Augustine JR. Circuitry and functional aspects of the insular lobe in primates including humans. Brain Research Reviews 1996;22:229–244. [PubMed: 8957561]

Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, et al. Voxel-based lesion-symptom mapping. Nature Neuroscience 2003;6:448–450.

Blank SC, Scott SK, Murphy K, Warburton E, Wise RJS. Speech production: Wernicke, Broca and beyond. Brain 2002;125:1829–1838. [PubMed: 12135973]

Brown R, McNeill D. The "tip of the tongue" phenomenon. Journal of Verbal Learning and Verbal Behavior 1966;5:325–337.

Burke DM, Locantore JK, Austin AA, Chae B. Cherry pit primes Brad Pitt: Homophone priming effects on young and older adults' production of proper names. Psychological Science 2004;15:164–170. [PubMed: 15016287]

Burke DM, MacKay DG. Memory, language and ageing. Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences 1997;352:1845–1856.

Burke DM, MacKay DG, Worthley JS, Wade E. On the tip of the tongue: What causes word finding failures in young and older adults? Journal of Memory and Language 1991;30:542–579.

Burke DM, Shafto MA. Aging and language production. Current Directions in Psychological Science 2004;13:21–24.

Burke, DM.; Shafto, MA. Language and aging. In: Craik, FIM.; Salthouse, TA., editors. The handbook of aging and cognition. New York: Psychology Press; 2008. p. 373-443.

Cereda C, Ghika J, Maeder P, Bogousslavsky J. Strokes restricted to the insular cortex. Neurology 2002;59:1950–1955. [PubMed: 12499489]

Christoff K, Prabhakaran V, Dorfman J, Zhao Z, Kroger JK, Holyoak KJ, et al. Rostrolateral prefrontal cortex involvement in relational integration during reasoning. Neuroimage 2001;14:1136–1149. [PubMed: 11697945]

Cohen JD, Botvinick MM, Carter CS. Anterior cingulate and prefrontal cortex: Who's in control? Nature Neuroscience 2000;3:421–423.

Cross ES, Burke DM. Do alternative names block young and older adults' retrieval of proper names? Brain and Language 2004;89:174–181. [PubMed: 15010248]

Dell GS. A spreading activation theory of retrieval in sentence production. Psychological Review 1986;93:283–321. [PubMed: 3749399]

Dogil G, Ackermann H, Grodd W, Haider H, Kamp H, Mayer J, et al. The speaking brain: A tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. Journal of Neurolinguistics 2002;15:59–90.

Dronkers NF. A new brain region for coordinating speech articulation. Nature 1996;384:159–161. [PubMed: 8906789]

Esposito G, Kirby BS, Van Horn JD, Ellmore TM, Berman KF. Context-dependent, neural system-specific neurophysiological concomitants of ageing: Mapping PET correlates during cognitive activation. Brain 1999;122:963–979. [PubMed: 10355679]

Etard O, Mellet E, Papathanassiou D, Benali K, Houde O, Mazoyer B, et al. Picture naming without Broca's and Wernicke's area. NeuroReport 2000;11:617–622. [PubMed: 10718324]

Evrard M. Ageing and lexical access to common and proper names in picture naming. Brain and Language 2002;81:174–179. [PubMed: 12081390]

Folstein MF, Folstein SE, McHugh PR. Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 1975;12:189–198. [PubMed: 1202204]

Gehring WJ, Knight RT. Prefrontal–cingulate interactions in action monitoring. Nature Neuroscience 2000;3:516–520.

Good CD, Johnsrude IS, Ashburner J, Henson RNA, Friston KJ, Frackowiak RSJ. A voxel-based morphometric study of ageing in 465 normal adult human brains. Neuroimage 2001;14:21–36. [PubMed: 11525331]

Habib M, Daquin G, Milandre L, Royere ML, Rey M, Lanteri A, et al. Mutism and auditory agnosia due to bilateral insular damage—Role of the insula in human communication. Neuropsychologia 1995;33:327–339. [PubMed: 7791999]

Harasty JA, Halliday GM, Xuereb J, Croot K, Bennett H, Hodges JR. Cortical degeneration associated with phonologic and semantic language impairments in AD. Neurology 2001;56:944–950. [PubMed: 11294934]

Hillis AE, Work M, Barker PB, Jacobs MA, Breese EL, Maurer K. Re-examining the brain regions crucial for orchestrating speech articulation. Brain 2004;127:1479–1487. [PubMed: 15090478]

Indefrey P, Brown CM, Hellwig F, Amunts K, Herzog H, Seitz RJ, et al. A neural correlate of syntactic encoding during speech production. Proceedings of the National Academy of Sciences, USA 2001;98:5933–5936.

Indefrey P, Levelt WJM. The spatial and temporal signatures of word production components. Cognition 2004;92:101–144. [PubMed: 15037128]

James LE, Burke DM. Phonological priming effects on word retrieval and tip-of-the-tongue experiences in young and older adults. Journal of Experimental Psychology: Learning, Memory, and Cognition 2000;26:1378–1391.

James, W. The principles of psychology. New York: H. Holt & Co; 1890.

Kaplan, EF.; Goodglass, H.; Weintraub, S. The Boston Naming Test. 2. Philadelphia, PA: Lea & Febiger; 1983.

Kikyo H, Ohki K, Sekihara K. Temporal characterization of memory retrieval processes: An fMRI study of the "tip of the tongue" phenomenon. European Journal of Neuroscience 2001;14:887–892. [PubMed: 11576194]

Kroger JK, Sabb FW, Fales CL, Bookheimer SY, Cohen MS, Holyoak KJ. Recruitment of anterior dorsolateral prefrontal cortex in human reasoning: A parametric study of relational complexity. Cerebral Cortex 2002;12:477–485. [PubMed: 11950765]

Levelt WJM. Spoken word production: A theory of lexical access. Proceedings of the National Academy of Sciences, USA 2001;98:13464–13471.

Lovelace EA, Twohig PT. Healthy older adults' perceptions of their memory functioning and use of mnemonics. Bulletin of the Psychonomic Society 1990;28:115–118.

MacKay, DG. The organization of perception and action: A theory for language and other cognitive skills. New York: Springer-Verlag; 1987.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 2003;19:1233–1239. [PubMed: 12880848]

Maril A, Simons JS, Weaver JJ, Schacter DL. Graded recall success: An event-related fMRI comparison of tip of the tongue and feeling of knowing. Neuroimage 2005;24:1130–1138. [PubMed: 15670690]

Maril A, Wagner AD, Schacter DL. On the tip of the tongue: An event-related fMRI study of semantic retrieval failure and cognitive conflict. Neuron 2001;31:653–660. [PubMed: 11545723]

Meyer AS, Bock K. The tip-of-the-tongue phenomenon: Blocking or partial activation? Memory & Cognition 1992;20:715–726.

Nelson, HE. National adult reading test. Windsor, UK: NFER-Nelson; 1982.

Oldfield RC. The assessment and analysis of handedness: The Edinburgh inventory. Neuropsychologia 1971;9:97–113. [PubMed: 5146491]

Prabhakaran V, Smith JAL, Desmond JE, Glover GH, Gabrieli JDE. Neural substrates of fluid reasoning: An fMRI study of neocortical activation during performance of the Raven's Progressive Matrices Test. Cognitive Psychology 1997;33:43–63. [PubMed: 9212721]

Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. Psychological Review 2000;107:460–499. [PubMed: 10941277]

Rastle KG, Burke DM. Priming the tip of the tongue: Effects of prior processing on word retrieval in young and older adults. Journal of Memory and Language 1996;35:586–605.

Raven, JC. Standard progressive matrices. London: H.K Lewis; 1958.

Raz N, Gunning-Dixon FM, Head D, Dupuis JH, Acker JD. Neuroanatomical correlates of cognitive aging: Evidence from structural magnetic resonance imaging. Neuropsychology 1998;12:95–114. [PubMed: 9460738]

Resnick SM, Pham DL, Kraut MA, Zonderman AB, Davatzikos C. Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain. Journal of Neuroscience 2003;23:3295–3301. [PubMed: 12716936]

Shipley, WC. Institute of living scale. Los Angeles: Western Psychological Services; 1946.

Sommer M, Koch MA, Paulus W, Weiller C, Büchel C. Disconnection of speech-relevant brain areas in persistent developmental stuttering. Lancet 2002;360:380–383. [PubMed: 12241779]

Sowell ER, Peterson BS, Thompson PM, Welcome SE, Henkenius AL, Toga AW. Mapping cortical change across the human life span. Nature Neuroscience 2003;6:309–315.

Stamatakis EA, Marslen-Wilson WD, Tyler LK, Fletcher PC. Cingulate control of fronto-temporal integration reflects linguistic demands: A three-way interaction in functional connectivity. Neuroimage 2005;28:115–121. [PubMed: 16023871]

Talairach, J.; Tournoux, P. Co-planar stereotaxic atlas of the human brain: 3-dimensional proportional system: An approach to cerebral imaging. Stuttgart: Georg Thieme Verlag; 1988.

Tonkonogy J, Goodglass H. Language function, foot of the third frontal gyrus, and rolandic operculum. Archives of Neurology 1981;38:486–490. [PubMed: 7247785]

Tyler LK, Marslen-Wilson WD, Stamatakis EA. Differentiating lexical form, meaning, and structure in the neural language system. Proceedings of the National Academy of Sciences, USA 2005a; 102:8375–8380.

Tyler LK, Marslen-Wilson WD, Stamatakis EA. Dissociating neuro-cognitive component processes: Voxel-based correlational methodology. Neuropsychologia 2005b;43:771–778. [PubMed: 15721189]

Wechsler, D. Manual for the Wechsler memory scale—Revised. San Antonio: The Psychological Corporation; 1987.

White KK, Abrams L. Does priming specific syllables during tip-of-the-tongue states facilitate word retrieval in older adults? Psychology and Aging 2002;17:226–235. [PubMed: 12061408]

Williams EJ. The comparison of regression variables. Journal of the Royal Statistical Society: Series B 1959;21:396–399.

Wise RJS, Greene J, Büchel C, Scott SK. Brain systems for word perception and articulation. Lancet 1999;353:1057–1061. [PubMed: 10199354]

Figure 1.
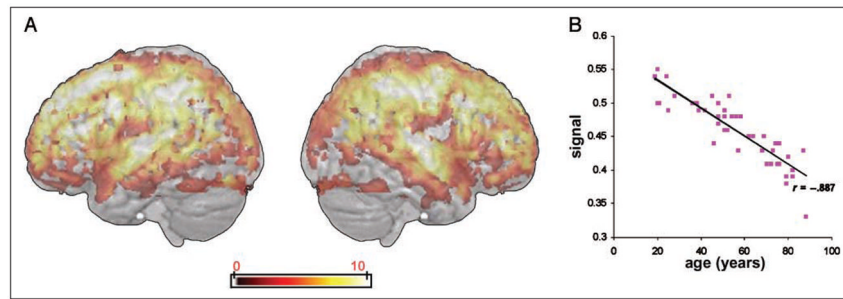Example of a picture and caption used on a single trial in the TOT task.

Figure 2.
Areas of significant negative correlation between gray matter density across the whole brain and age in years. (A) Cortical areas where gray matter density correlates negatively with participant age are shown on a three-dimensional rendered spatially normalized brain. The color bar indicates the strength of correlation (voxel-level *T* values). (B) Scatterplot showing the relationship between mean gray matter density across the whole brain and age of participants in years.
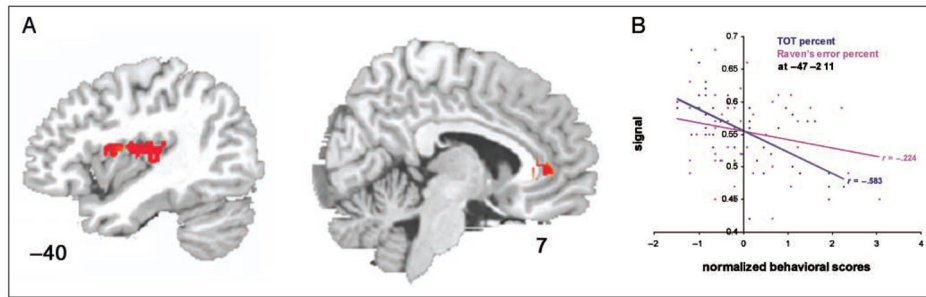
Figure 3.
Areas of significant negative correlation between gray matter density across the whole brain and proportion of TOT experiences. (A) Sagittal slices from a T1-weighted spatially normalized image of a single subject, showing clusters of significant negative correlation between proportion of TOTs and gray matter density when RPM scores are factored out (statistical peak, $x = -47$, $y = -2$, $z = 11$; Z-score = 4.31). Talairach x values are in mm. (B) Correlation coefficients show the relationship between gray matter density and normalized TOT and RPM scores at the most significant voxel in the whole-brain analysis ($x = -47$, $y = -2$, $z = 11$; Z-score = 4.07). A Williams (1959) test shows that the two correlations are statistically different, $p < .05$.
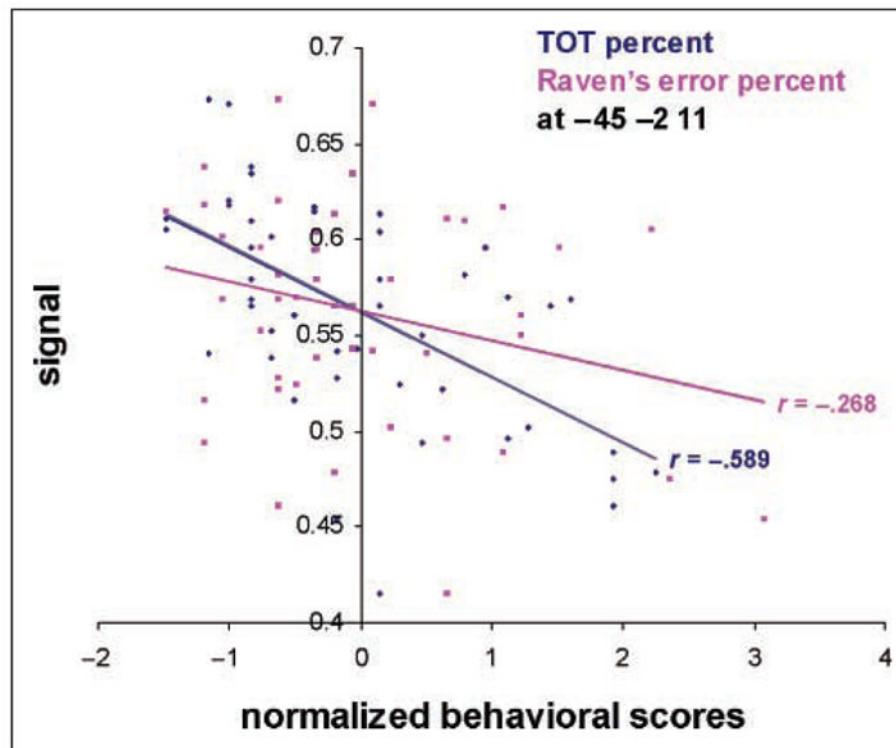
Figure 4.
Scatterplot showing the relationship between gray matter density in the left insula and normalized TOT and RPM scores. Correlations are for the most significant voxel within the left insula ($x = -45$, $y = -2$, $z = 11$; Z-score = 4.33). A Williams (1959) test shows that the two correlations are statistically different, $p < .05$ (one-tailed).
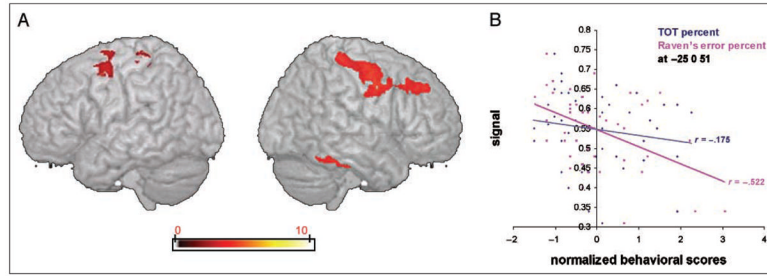
Figure 5.
Areas of significant negative correlation between gray matter density across the whole brain and proportion of RPM errors. (A) Areas of significant correlation between gray matter density and RPM across the whole brain are shown on a three-dimensional rendered spatially normalized brain (statistical peak, $x = -25$, $y = 0$, $z = 51$; Z-score = 4.19). The color bar indicates the strength of correlation (voxel level T values). (B) Correlation coefficients showing the relationship between gray matter density and normalized TOT and RPM scores at the most significant voxel in the whole brain analysis ($x = -25$, $y = 0$, $z = 51$; Z-score = 4.19). A Williams (1959) test shows that the two correlations are statistically different, $p < .05$ (one-tailed).