Neuron
Article

Lichtheim 2: Synthesizing Aphasia and the Neural Basis of Language in a Neurocomputational Model of the Dual Dorsal-Ventral Language Pathways

Taiji Ueno,[1] Satoru Saito,[1,2] Timothy T. Rogers,[3] and Matthew A. Lambon Ralph[1,*]
[1]Neuroscience and Aphasia Research Unit, School of Psychological Sciences, University of Manchester, Manchester M13 9PL, UK
[2]Kyoto University, Kyoto, 606-8501, Japan
[3]University of Wisconsin-Madison, Madison, WI 53706-1611, USA
*Correspondence: matt.lambon-ralph@manchester.ac.uk
DOI 10.1016/j.neuron.2011.09.013

SUMMARY

Traditional neurological models of language were based on a single neural pathway (the dorsal pathway underpinned by the arcuate fasciculus). Contemporary neuroscience indicates that anterior temporal regions and the "ventral" language pathway also make a significant contribution, yet there is no computationally-implemented model of the dual pathway, nor any synthesis of normal and aphasic behavior. The "Lichtheim 2" model was implemented by developing a new variety of computational model which reproduces and explains normal and patient data but also incorporates neuroanatomical information into its architecture. By bridging the "mind-brain" gap in this way, the resultant "neurocomputational" model provides a unique opportunity to explore the relationship between lesion location and behavioral deficits, and to provide a platform for simulating functional neuroimaging data.

INTRODUCTION

The pioneering 19th century neurologists demonstrated that language is supported by a distributed network of cortical regions. Lichtheim (1885) assimilated these ideas into a model that implicated left-hemisphere perisylvian and prefrontal regions, and proposed that these communicate with one another in order to generate language function. Lichtheim's model accounted for the main forms of language impairment following brain damage, explaining why lesions in different brain regions might produce different aphasic syndromes. One hundred twenty-five years after it was proposed, Lichtheim's model remains the main organizing framework for thinking about the neural basis of language and its pathologies for many researchers and clinicians.

Despite its considerable and long-standing success, early critics noted the theory's lack of specificity regarding the functions computed by the different cortical regions. More recently it has become clear that a wealth of information needs to be incorporated including contemporary neuroscience data about the functional and structural anatomy of the language system. The current paper offers a new proposal about the neurocomputation of language that is similar in spirit to Lichtheim's enterprise but that incorporates new facts about the structure and function underpinning language. Specifically, we propose that single-word comprehension, production (speaking/naming) and repetition are supported by the interactive contributions of the dorsal and ventral language pathways.

Our model draws on important and influential contributions from prior computational models of language and short-term memory (Botvinick and Plaut, 2006; Dell et al., 1997; Dilkina et al., 2008, 2010; McClelland et al., 1989; Nozari et al., 2010; Plaut and Kello, 1999; Plaut et al., 1996; Rogers et al., 2004; Seidenberg and McClelland, 1989; Welbourne and Lambon Ralph, 2007; Woollams et al., 2007). We reconfigured the architectures employed in these purely computational models to better reflect our current state of knowledge about the actual neuroanatomy of the language system. With the resulting neuroanatomically-constrained computational model, we show how both classical and progressive forms of aphasia arise within this architecture and how it explains well-established lesion-symptom correlations for each form. We further demonstrate how the incorporation of neuroanatomy within an explicit neurocomputational formalism addresses the shortcomings of Lichtheim-era models. First, quantitative analysis of internal representations developed by the model allows the theorist to specify the nature of the functions computed by each brain region and to relate these to empirical observations from functional neuroimaging. Second, simulations of plasticity-related recovery offer explicit and testable hypotheses about the partial spontaneous recovery observed in many patients post damage.

The classic neuroanatomical model of language is dominated by the "dorsal" language pathway, in which perisylvian language areas, such as posterior superior temporal gyrus (pSTG), inferior supramarginal gyrus (iSMG), and inferior frontal gyrus (iFG), are connected via the arcuate fasciculus (AF) (Eggert, 1977; Geschwind, 1970; Lichtheim, 1885). The recent neuroscience literature (and quite possibly Wernicke himself) (Weiller et al., 2011) has gone beyond those areas and single pathway to suggest key roles for rostral temporal regions connected via a "ventral" pathway (including the middle longitudinal fasciculus, extreme capsule [EmC] and possibly the uncinate fasciculus [UF])(Hickok

and Poeppel, 2000, 2007; Parker et al., 2005; Rauschecker and Scott, 2009; Saur et al., 2008; Scott et al., 2000). Most proposals suggest that there is at least some specialization of function in each pathway, with the dorsal pathway capturing the sound-to-motor statistical structure in a given language (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009) and being the principal source of repetition deficits in aphasic patients (Fridriksson et al., 2010). In contrast, the ventral pathway is associated with lexical-semantic influences on word repetition (Nozari et al., 2010) and relatedly, this pathway is crucial in the extraction of meaning from the acoustic-phonological input (Hickok and Poeppel, 2007; Scott et al., 2000), where it converges—in anterior temporal regions—with conceptual knowledge formed from the integration of information from other sensory inputs (Lambon Ralph et al., 2010; Patterson et al., 2007; Visser and Lambon Ralph, 2011). Although not a part of the classical neurological models of language, there is now considerable evidence for the importance of this region and pathway from functional neuroimaging (Binder et al., 2011; Binney et al., 2010; Scott et al., 2000; Sharp et al., 2004), rTMS (Holland and Lambon Ralph, 2010; Pobric et al., 2007), patients (Patterson et al., 2007; Schwartz et al., 2009), and direct white-matter and cortical stimulation studies (Duffau et al., 2009; Matsumoto et al., 2004).

The implemented dual-pathway model permits a formal consideration of how this region and its associated processing fit into the larger language network, and also provides a crucial test of the face validity of the underlying computational and neuroanatomical assumptions. The model also licenses a unique opportunity to probe the computations performed within each "region," as well as their interactions directly. By marrying computation and neuroanatomy in this way, the resultant neurocomputational model provides a foundation for understanding the neural basis of aphasia as well as a framework for simulating functional neuroimaging results (both of which require a bridge between anatomy and behavior).

Aims of the Current Study

Assessing the Face Validity of the "Lichtheim 2" Model

Using techniques and methods from prior work, we reformulated the architecture of purely computational models of language in order to mirror the dual pathway neuroanatomy, more faithfully. We tested the resulting "neurocomputational" model for its ability to synthesize normal language behaviors (speaking/naming, comprehending, and repeating). With a successful model in place, we were then able to probe each "neuroanatomical" element of the framework in order to undercover its function and the interaction between the two pathways.

Understanding Aphasia in Lichtheim 2

The dual-pathway hypothesis has emerged in the context of contemporary tractography, functional neuroimaging, and aphasiological data (Hickok and Poeppel, 2007; Nozari et al., 2010; Parker et al., 2005; Rauschecker and Scott, 2009) whereas the classic models of language were primarily aimed at explaining (though not synthesizing) different types of chronic aphasia (Eggert, 1977; Geschwind, 1970; Lichtheim, 1885). We explored, therefore, how chronic (stroke-related) and progressive (semantic dementia) forms of aphasia emerge after damage to the Lichtheim 2 neurocomputational model. In addition, we also addressed the emerging view that chronic patient performance reflects the combination of damage and partial recovery processes, which might follow from a reweighting of neural connections in order to reoptimize the remaining computational capacity (Lambon Ralph, 2010; Leff et al., 2002; Sharp et al., 2010; Welbourne and Lambon Ralph, 2007).

Understanding Contemporary Neuroscience Findings

The rise of sophisticated structural and functional neuroimaging has spawned a wealth of new information about (1), the computations associated with different parts of the language network; and (2), the nature of patients' impaired language function after different locations of damage. We tested the model's ability to capture and explain a high-profile example of each type: (1), the acoustic/phonological-to-semantic transformation of information along the ventral, rostral pathway (Griffiths, 2002; Scott et al., 2000) by undertaking an analysis of the changing similarity-structure encoded at different points along the ventral pathway; and (2), by assessing the rate of semantic speaking/naming errors after different lesion locations, we tested whether the peak semantic error rate follows from damage to the anterior STG as demonstrated in a recent voxel-symptom lesion mapping study (Schwartz et al., 2009). In turn, by probing and understanding the nature of the information coding in this region, the model provided insights about why lesions in this, but not other locations, generate a maximal number of semantic errors.

Understanding the Significance of the Dual Pathways

Although there is clear and emerging evidence of dual language pathways in the human brain, neurocomputational models allow us to test the functioning of different possible architectures (for a parallel computational comparison with respect to naming and repetition in aphasia see Nozari et al. [2010]). By implementing a single-pathway architecture and comparing it with the dual-pathway model, we were able to explore why it might be beneficial for the real brain to utilize dual, interactive pathways for language.

RESULTS

Figure 1 shows the neuroanatomically-constrained architecture of the dual-pathway model (see Experimental Procedures; see Figure S1 available online for further details). This includes a dorsal pathway (primary auditory ↔ inferior supramarginal gyrus ↔ insular-motor cortex: underpinned by the arcuate fasciculus) and a ventral pathway (primary auditory ↔ mid-STG ↔ anterior-STG ↔ opercularis-triangularis ↔ insular-motor cortex: underpinned by the middle longitudinal fasciculus and the extreme capsule). Via a connection to the aSTG, we also incorporated the ventral anterior temporal region (vATL), which has been shown by functional neuroimaging and neuropsychological studies to be crucial to verbal and nonverbal comprehension (Binney et al., 2010; Mion et al., 2010; Visser and Lambon Ralph, 2011) . The model was trained to speak/name, repeat and comprehend a set of 1710 multisyllabic Japanese words (incorporating time-varying input and outputs, as well as other computationally-challenging characteristics of language; see Experimental Procedures). Figure 2 shows the developmental trajectory of the network on these tasks as a function of

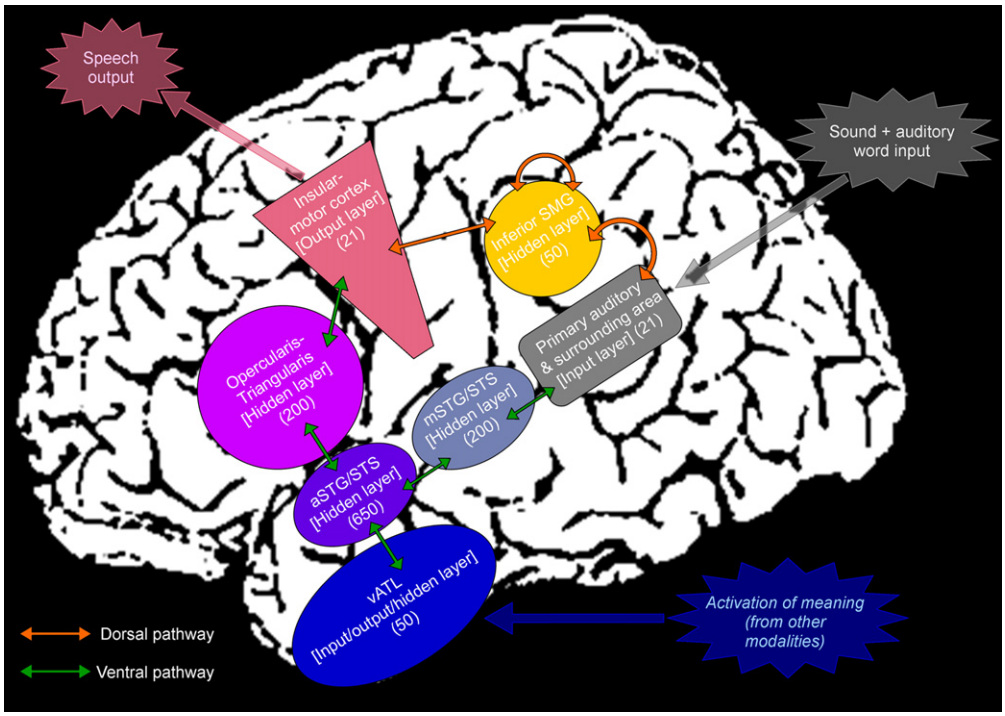

Figure 1. Lichtheim 2—the Implemented Neuroanatomically-Constrained Dual-Pathway Language Model
See also Figure S1.

word-frequency. Like children, the model demonstrated different acquisition functions for each task with repetition preceding comprehension and comprehension preceding speaking/naming. In its trained, "adult" state, the model was able to repeat all of the words from the training set, other untrained real words (i.e., real Japanese words that were not included in the training set) and a set of nonwords (i.e., legitimate Japanese phonemic nonword sequences which, inevitably, had lower phonotactic and bi-mora frequency than the untrained real words), with performance comparing closely to human data (see Figure S2). In summary, the implemented dorsal-ventral neurocomputational model proved to be a fully-functional model compatible with adult and children's language performance.

Synthesizing Aphasic Patient Data

Figure 3 summarizes the effect of simulated lesions to the different regions (representational layers) in the model. Performance was assessed with 60 high frequency and 60 low frequency words (see Supplemental Experimental Procedures for details). Fifteen levels of damage severity were simulated in each region. Most forms of lesion or atrophy include damage to both gray matter and the underlying white matter. Accordingly, simulated damage included both the addition of noise to the unit outputs as an analog of gray matter pathology (ranging from 0.01 to 0.15 in equal intervals) and removal of the incoming links to the damaged layer as an analog of white matter damage (ranging from 0.5% to 7.5% in equal intervals).

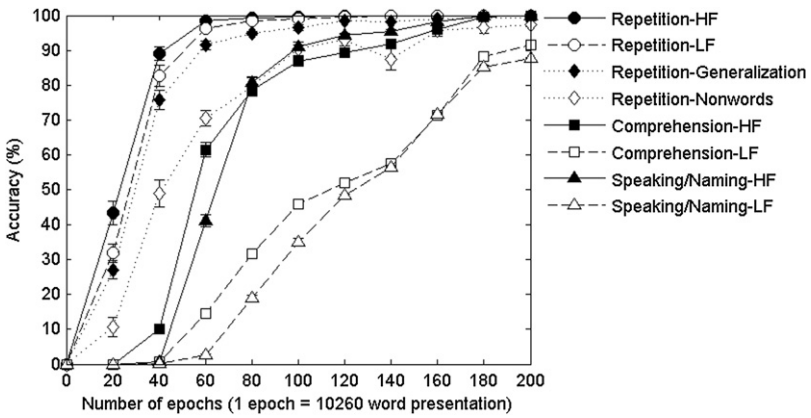

Figure 2. Accuracy of the Network on Three Language Tasks as a Function of Learning
The y axis error bars show the standard errors across five simulations. Generalization of repetition was tested using both untrained items (real Japanese words) as well as novel tri-mora items (Japanese nonwords) for direct comparison with human participants. See also Figure S2.

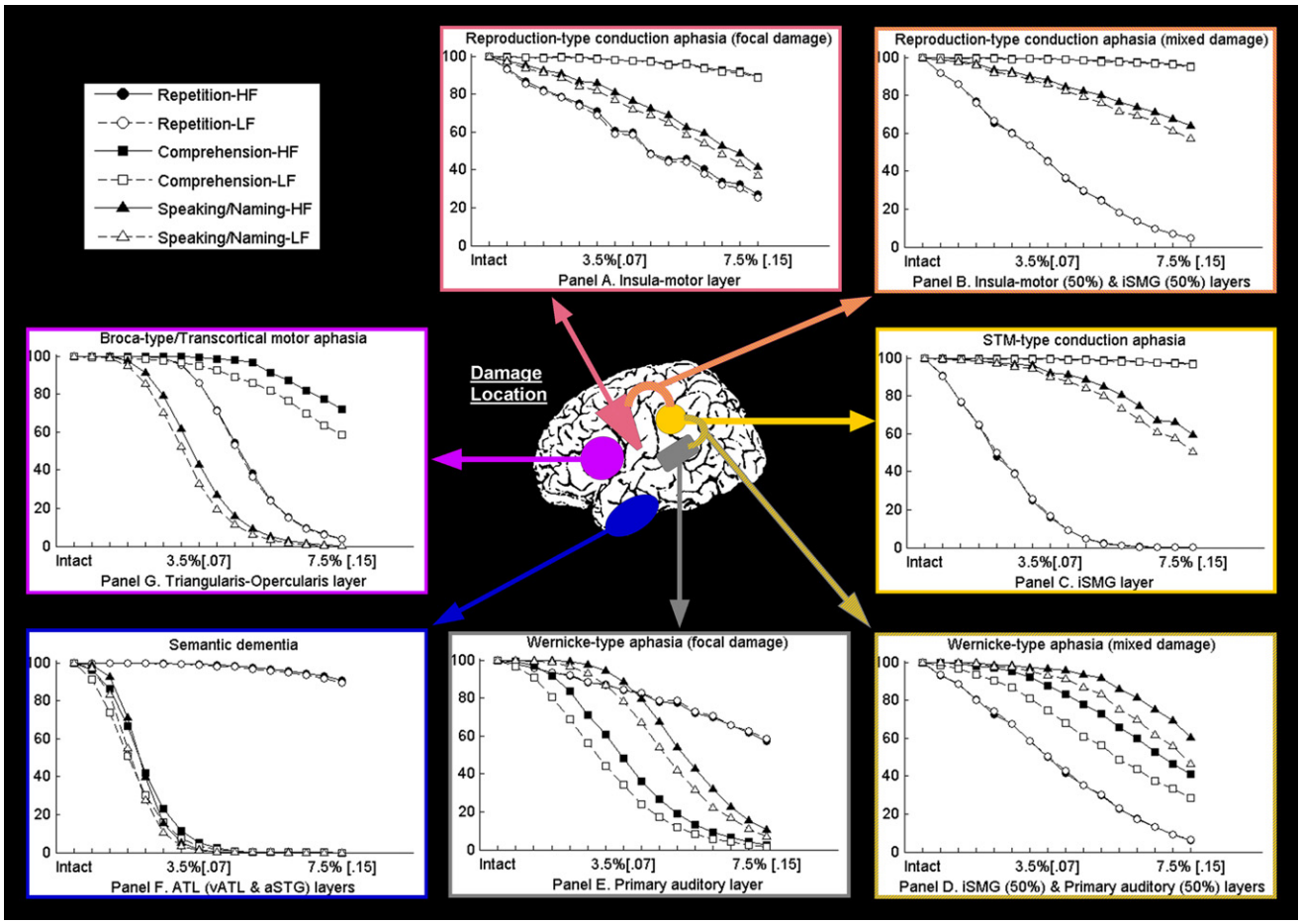

Figure 3. Synthesis of Classic (Stroke) and Progressive Aphasia as a Result of Different Lesion Locations and Severities

The x axis shows the proportion (%) of the incoming links removed and the range of the noise (bracket) over the output of the damaged layer. "Intact" means the performance of the model before lesioning.

Several classic (stroke) and progressive aphasias were synthesized. Figures 3A–3E summarize the impact of lesions to the posterior perisylvian region (with the lesion focus shifting in a caudal-inferior, "clockwise" fashion). Damage to the insular-motor layer led to reproduction conduction aphasia (impaired spoken output with preserved comprehension) and a similar pattern was generated when the lesion covered the insular-motor and iSMG layers (Figures 3A and 3B). The selective pattern of STM/repetition conduction aphasia (Shallice and Warrington, 1977) followed from isolated iSMG lesioning (Figure 3C) although with greater levels of damage both speaking/naming and repetition were compromised. Taken together, these simulations might explain the apparent rarity of STM/repetition over reproduction conduction aphasia, in that repetition-selective deficits only arise in the context of isolated and mild lesions to the iSMG layer. Overall, these simulations mirror the association between conduction aphasia and damage to the dorsal pathway observed in real patients (Fridriksson et al., 2010; Geschwind, 1965; Hillis, 2007). Wernicke's aphasia (severely impaired comprehension combined with moderate-to-severe impairments of speaking/naming and repetition) is associated with damage centered on the pSTG and surrounding region (Hillis, 2007). Damage to the corresponding part of the model (the acoustic-phonological input layer ± additional damage to the iSMG) resulted in exactly this behavioral pattern (Figures 3D and 3E). In contrast, lesions in iFG are known to result in a Broca-type/transcortical motor aphasia (Hillis, 2007) characterized, in the context of single-word processing, in terms of relatively good comprehension, impaired repetition and severely affected speaking/naming. Exactly this pattern followed in the model after damage to the corresponding region (the triangularis-opercularis layer; see Figure 3G). The final target was semantic dementia, epitomized by intact repetition with severely impaired comprehension and speaking/naming, especially for low-frequency words (Hodges et al., 1992; Jefferies et al., 2009; Lambon Ralph et al., 1998) in the context of atrophy focused on the inferolateral and polar aspects of the anterior temporal lobe (Galton et al., 2001; Hodges et al., 1992). Again, the model demonstrated this specific symptom combination following damage to the ATL components (vATL and

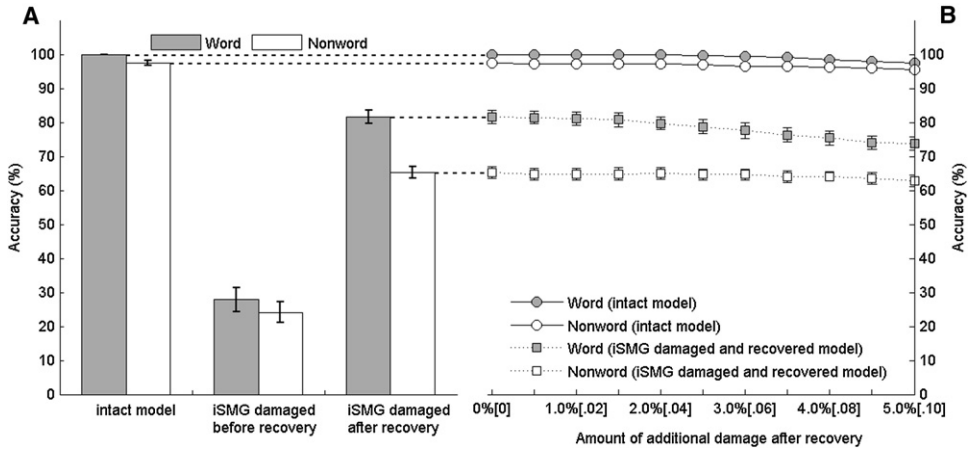

Figure 4. Simulation of Recovery Postdamage and the Changing Division of Labor

(A) Repetition accuracy after iSMG damage as a function of lexicality and recovery.

(B) Effect of subsequent ATL (diagnostic ventral pathway) damage on repetition. See main text for explanation.

aSTG; Figure 3F). For a formal comparison of the size of the word-frequency effect in the model versus real SD patients, we extracted a subset of words in order to match the size of the frequency manipulation used by Jefferies et al. (2009) (Cohen's *d* for HF-LF difference = 1.61 in our materials, and *d* = 1.64 in Jefferies et al. [2009]). With this test set, the HF-LF difference in comprehension accuracy of our model was 19.49% (1.5% weight removal and noise range = 0.03), which was close to the mean HF-LF difference in synonym judgment accuracy of the real SD patients in Jefferies et al. (2009) (18.52% in the high imageability condition).

In summary, the neurocomputational dual pathway model was able not only to synthesize the different symptom complexes of classic (stroke-related) and progressive aphasias but also to capture the link between each aphasia type and the different underlying location of damage. These lesion simulations also provide key insights about the underlying process of each language pathway. Rather than being equipotential for all tasks, the two pathways formed a partial division of labor such that damage to the dorsal pathway (Figures 3A–3C) had a larger impact on repetition while damage to the layers in the ventral pathway (Figures 3E and 3F) had a greater impact on comprehension and speaking/naming.

Next, an emerging view is that chronic patient performance reflects the combination of damage and partial recovery processes (Lambon Ralph, 2010; Leff et al., 2002; Sharp et al., 2010; Welbourne and Lambon Ralph, 2007). Thus, to capture and explore the basis of the partial recovery observed in aphasic patients in the year or more after their stroke, the damaged model was allowed to "recover" by reexposing it to the three language tasks and updating its remaining weight structure (using the same iterative weight-adjustment algorithm as per its development) (Welbourne and Lambon Ralph, 2007). For brevity and given the considerable computational demands associated with this kind of recovery-based simulation, we focused on one worked example in detail: iSMG damage leading to repetition conduction aphasia (Figure 3C: 1.0% removal of the incoming links; output noise [range = 0.1]; see Supplemental Experimental Procedures for details). The principal pattern of conduction aphasia (impaired repetition, mildly impaired naming and preserved comprehension) remained post recovery. In addition, there was a quantitative change in the size of the lexicality effect on repetition performance. Figure 4A shows word and nonword repetition accuracy pre- and postrecovery (20 epochs of language exposure and weight update). Like human adults, a small lexicality effect was observed in the intact model ($t(4) = 3.81$, $p = 0.019$, Cohen's $d = 1.90$). Immediately after damage, both word and nonword repetition was affected to an equal extent (the lexicality effect remained but was unchanged: $t(4) = 2.92$, $p = 0.043$, $d = 1.46$). Following language re-exposure not only was there partial recovery of repetition overall but also a much stronger lexicality effect emerged ($t(4) = 7.36$, $p = 0.002$, $d = 3.68$) of the type observed in aphasic individuals (Crisp and Lambon Ralph, 2006).

Diagnostic simulations (additional damage to probe the functioning of a region pre- and postrecovery) revealed that these recovery-related phenomena were underpinned in part by a shift in the division of labor (Lambon Ralph, 2010; Welbourne and Lambon Ralph, 2007) between the pathways, with an increased role of the ventral pathway in repetition. Figure 4B summarizes the effect of increasing diagnostic damage to the ATL (vATL and aSTG layers) on the partially-recovered model. A three-way ANOVA with factors of lexicality, model-status (intact versus recovered model), and ATL-lesion severity, revealed a significant three-way interaction ($F(10, 40) = 7.78$, $p < 0.001$). The lexicality × ATL-lesion severity interaction was not significant before recovery ($F(10, 40) = 1.73$, $p = 0.11$) but was significant after recovery ($F(10, 40) = 12.44$, $p < 0.001$). Moreover, the recovery × ATL-lesion severity interaction was significant for words ($F(10, 40) = 9.49$, $p < 0.001$) but not significant for nonwords ($F(10, 40) < 1$). In summary, while maintaining the core features of conduction aphasia, the network reoptimized repetition performance, in part, by reallocating the intact resource from the ventral pathway. Word repetition benefits the most from this changed

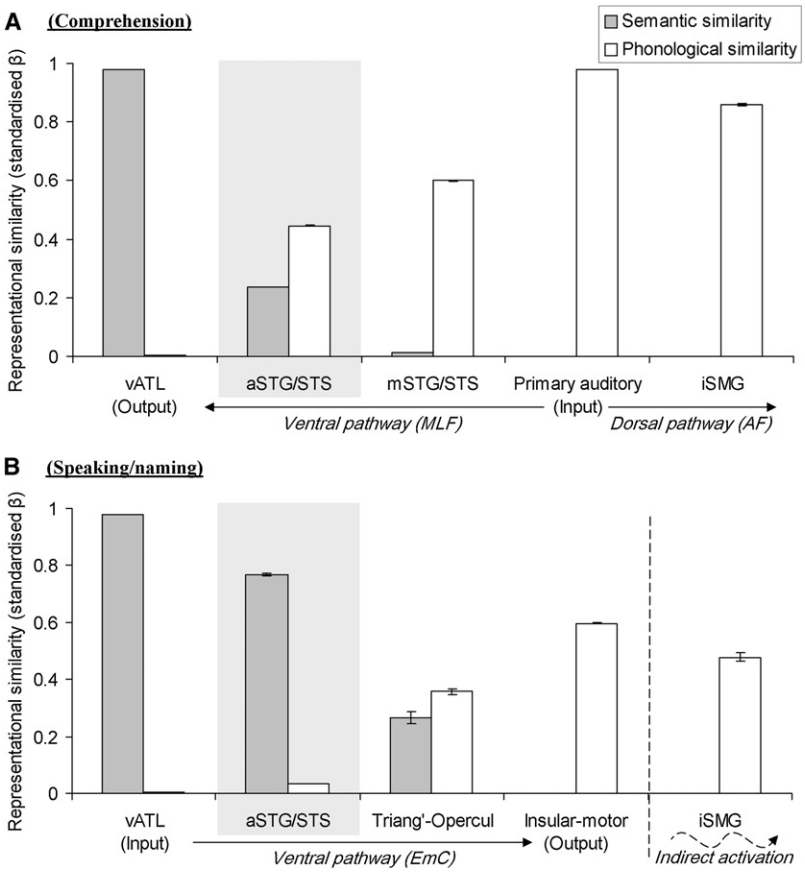

Figure 5. Analysis of the Representational Similarity Structure in Each Layer

Data are presented from the comprehension trials (A) versus speaking/naming trials (B). The similarity of the representations at each layer was predicted from pure measures of semantic and phonological similarity. High standardized regression coefficients (β) denote a strong relationship. Trian′-Opercul = triangularis-opercularis.

division of labor because the ventral pathway is intrinsic to the processing of meaning (which, by definition, nonwords do not have). These results complement previous explorations of aphasic repetition and naming performance with respect to dual versus single language pathways (Nozari et al., 2010).

Simulation of Contemporary Neuroscience Findings

Modern neuroimaging techniques provide important information beyond that offered by patient studies alone. This includes the ability to probe the function of a region and how this changes across neighboring areas. The implementation of a neurocomputational model licenses an investigation of these types of contemporary neuroscience data. For example, Scott et al. (2000) demonstrated an acoustic/phonological-to-semantic rostral shift in function along the ventral language pathway. We simulated these specific results by probing the similarity structure of the representations formed across different components of the model. The rationale here is that if a layer is responsible for semantic processing, for example, then semantically-related items should be similarly represented in that layer.

In the first analysis, we probed the successive layers of the ventral and dorsal pathways after the presentation of an acoustic-phonological input. A series of multiple regressions was used to probe the similarity of the activation observed at each layer by using pure semantic and phonological similarity as predictor variables (see Experimental Procedures and Supplemental Experimental Procedures for details). Figure 5A shows the standardized β values which summarize how strongly phonological and semantic similarity predicted the observed activation similarity at each layer. Unsurprisingly, given that they had been trained to do so, the vATL layer captured semantic similarity and the primary auditory layer (input) captured phonological similarity. These two regions/layers were included as semantic and phonological references against which the other layers (the representations of which had not been prespecified) could be compared. The iSMG layer in the dorsal pathway was strongly sensitive to phonological similarity, consistent with the notion (and simulation, above) that this pathway is crucial for repetition/mimicry (Rauschecker and Scott, 2009). In the ventral pathway, the two intermediate layers before the vATL exhibited the graded phonological-to-semantic rostral progression observed in humans (Scott et al., 2000). Specifically, the first layer (mSTG) primarily captured acoustic-phonological similarity while the second component (aSTG) was more sensitive to semantic and less to phonological similarity.

The same analysis was conducted on the output side of the ventral pathway (Figure 5B). The activation patterns were measured at the third time tick of speaking/naming (see Supplemental Experimental Procedures). Again there was a gradual shift in the type of similarity structure encoded across the successive layers (vATL to the insula-motor output layer: left to right in Figure 5B), becoming increasingly sensitive to phonological similarity and less so to semantic similarity.

Qualitative Nature of Impairments

While the pattern of behavioral dissociations varies according to the location of brain damage, the qualitative nature of impairments can also change. For example, recent voxel-symptom lesion mapping studies have demonstrated significant variation in the rate of semantic speaking/naming errors according to the location of stroke-related damage, peaking in the aSTG (Schwartz et al., 2009). The rate of semantic errors produced by the model was compared after various levels of damage to each of its internal layers (to permit a fair comparison, the level of damage for each was titrated to equate overall speaking/naming accuracy, and lesioning was repeated ten times with different random seeds to avoid idiosyncratic results [Figure 6A]; see Supplemental Experimental Procedures for further

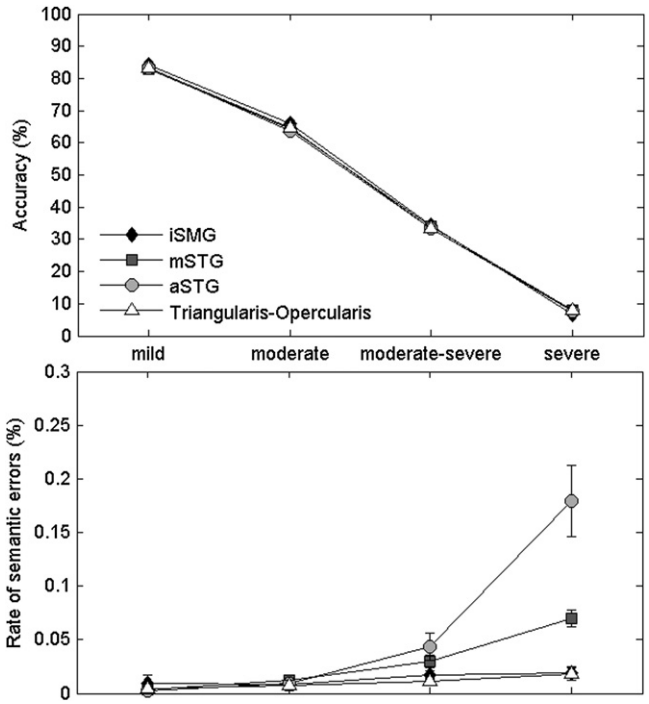

Figure 6. Speaking/Naming Accuracy and Rate of Semantic Errors as a Function of Damage Location

Lesion severity was titrated across the different layers in order to match speaking/naming accuracy irrespective of lesion location (A). Both accuracy and semantic errors (B) are expressed as a proportion of the total number of presented items (166).

methodological details). Figure 6B (bottom) shows the rate of semantic errors as a function of the location and degree of damage. A 4 (place of damage) × 4 (severity level) ANOVA revealed a significant interaction ($F(9, 81) = 19.27$, $p < 0.001$). Increasing damage in aSTG significantly augmented the rate of errors ($F(3, 27) = 33.26$, $p < 0.001$). This pattern, albeit less pronounced, was also found for mSTG ($F(3, 27) = 40.773$, $p < 0.001$), but not for iSMG or opercularis-triangularis (Fs < 1). In parallel to the patient data, these simulations revealed that the rate of semantic errors was most pronounced after aSTG simulated damage (aSTG versus mSTG: $F(1, 9) = 10.82$, $p = 0.009$). Importantly, like the original patient study, these simulation analysis outcomes remained, even after the comprehension accuracy was controlled (ANCOVA).

To explain their results, Schwartz et al. (2009) proposed that aSTG mediates lexical access for speech production. This raises a conundrum, however, in that the same region is associated with both verbal and nonverbal auditory comprehension in patient, rTMS, and functional neuroimaging literature (Patterson et al., 2007; Pobric et al., 2007; Scott et al., 2000; Visser and Lambon Ralph, 2011). Using the regression-based method (see above), we probed the functioning of the aSTG layer of the model across tasks (comparing the aSTG-associated β values [highlighted with a light-gray box] in Figure 5A [comprehension] versus Figure 5B [speaking/naming]). Given the structure of the model and the lesion simulations summarized above, it is clear that the region is critical both in auditory comprehension and speaking/naming. In addition, these analyses demonstrated that the similarity structure coded at the same layer is dependent on the task, such that it is much more semantically-influenced and less phonologically-bound during speaking/naming than comprehension. When any computational system is impaired, the error types reflect the underlying similarity structure, and so it is unsurprising that the paraphasias following aSTG damage are primarily semantic in nature. In short, the dual-pathway model is able to capture not only the localization of different language functions across regions (as indicated by neuropsychological dissociations, rTMS, and functional imaging) but also the qualitative variation of patient performance.

Significance of Dual Pathways

Finally, although there is clear and emerging evidence of a dual language pathway in the human brain, the neurocomputational models allow us to test the functioning of different possible architectures (see also Nozari et al., 2010). Accordingly, we compared the dual-pathway model to a "ventral only" architecture that could, in principle, achieve the same three language activities (comprehension, repetition, and speaking/naming). The architecture of the ventral-only model (Figure 7A) differed from the standard model in the absence of the iSMG layer and its associated connectivity (the dashed gray arrows and layer). The ventral pathway (black solid arrows/layers) and all training parameters were identical with those of the standard model. Figure 7B summarizes the learning curves of the ventral-only model. Two major deviations from human behavior are immediately obvious from Figure 7: (1), repetition lagged behind comprehension and speaking/naming, rather than in advance of it as in the developmental profile of children; and (2), nonword repetition and generalization accuracy (diamond markers) were nonexistent (along the x axis). In effect, it would appear that the ventral pathway accomplished repetition (of words alone) solely on the basis of understand-then-name the acoustic-phonological input and thus, unlike real humans, had no ability to deal appropriately with novel stimuli (see also Figure S3 for another control simulation).

In general, when all tasks are supported by the same single pathway, the model will struggle to acquire the two types of mapping that underpin comprehension, speech/naming and nonword repetition. The relationship between speech sounds or speech gestures and semantics is essentially arbitrary. A system that learns to map from speech sounds to semantics (in comprehension) and from semantics to phonotactics (in production) will thus acquire intermediating representations that discard the shared structure that exists between speech sounds and phonotactics. In contrast, a model that adopts two pathways—one that involves semantics and one that does not—will be capable of mastering both the arbitrary mappings needed to support comprehension and production, and the systematic mappings existing between speech sounds and articulatory gestures. Knowledge of the systematic mappings supports both repetition and pronunciation of unknown words—skills that are likely to be critical for language acquisition during infancy and for the rapid deployment of novel forms in adulthood.

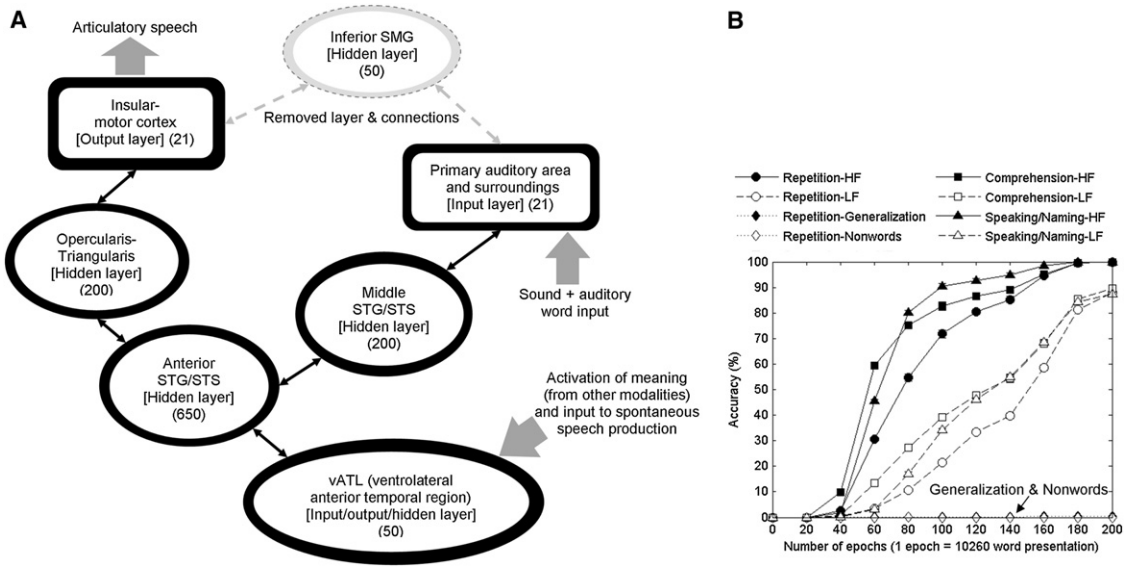

Figure 7. The Architecture of the "Ventral-Only" Model and Its Accuracy in Three Language Tasks as a Function of Learning

The black-solid layers/connections in the architecture (A) remained the same as the standard model whereas the gray-dashed layer/connections were removed. In (B), the diamond-marker lines (repetition of nontrained words and nontrained nonwords) were at floor. See also Figure S3 for another comparative simulation.

DISCUSSION

By fusing neuroanatomical information and computational modeling, the resultant neurocomputational framework was able to simulate normal and aphasic language profiles, as well as various forms of contemporary neuroscience data. Past computational models have generated critical insights about cognitive language processes and impaired function in neuropsychological patients but have made only limited contact with structural and functional neuroimaging data. Likewise, neuropsychology, neuroimaging, and other cognitive neuroscience methods provide crucial analytics for probing brain function but cannot offer a synthesis of normal and impaired function. The current neurocomputational model provides a foundation for the fusion of neuroanatomy and computation in the domain of language function. While future endeavors will be able to incorporate other brain regions, pathways, and behavioral data, the current simulations shed light on a range of core classical aphasiological data and contemporary neuroscience findings. More specifically, the model represents a neuroanatomically constrained implementation and validity test of the dual pathways framework, thus extending the classic Lichtheim model (itself never computationally implemented). As well as offering an explanation of key behavioral results, the Lichtheim 2 model provides an opportunity to explore the contribution of each element. These investigations highlighted three key phenomena that are summarized briefly below.

Anatomical Connectivity and the Nature of the Computational Mapping Determine Functional Connectivity

Except for the three peripheral layers, the model was free to develop its own representations and processing in each pathway. Given its proximity to the semantic-based representations of the vATL, the functioning of the ventral pathway becomes dominated by the input ↔ semantic ↔ output mappings which are doubly computationally challenging in that the mappings are both arbitrary in form and require transforming between time-varying (acoustic-phonology-motor) and time-invariant (semantic) representations (see Experimental Procedures). In turn, the same partial division of labor means that the dorsal pathway becomes somewhat independent of semantic influences and thus is better placed to encode the statistical regularities between acoustic-phonological and phonological-motor systems—such that this information can be generalized to novel forms (i.e., the model can repeat nonwords). Indeed, an additional simulation (Figure 7) indicated that it is difficult for a single (ventral) pathway to capture all these functions simultaneously because repetition becomes dominated by semantic influences so that the system is incapable of repeating novel word forms (which by definition have no meaning). In short, a dual-pathway architecture permits the parallel but separate extraction of mappings from the time-sensitive phonological-sound representations to other systematic time-varying representations (articulation) and to unsystematic time-invariant semantic statistical structures. The joint learning of both mappings in a single-pathway appears to be difficult or impossible. The corollary of these computational insights is that the double dissociations between certain types of aphasia (e.g., conduction aphasia—impaired repetition versus semantic dementia—impaired comprehension and speaking/naming) reflect these same divisions of labor in the human brain.

Division of Labor Is Plastic and Can Shift in the Context of Recovery

The simulations also suggest that the division of labor between the two pathways is not absolute or mutually exclusive. The

two pathways work together to deliver adult language performance (and aphasic naming and repetition abilities; see Nozari et al. [2010]). This division of labor represents one solution for an intact, fully-resourced computational model. The solution is not fixed, however, and following damage, processing can be re-optimized both within and across the two pathways, thereby mimicking spontaneous recovery observed post stroke (Lambon Ralph, 2010; Leff et al., 2002; Sharp et al., 2010; Welbourne and Lambon Ralph, 2007). These simulations suggest that this recovery sometimes comes at the cost of other functions (e.g., more of the computation underpinning repetition can be taken up by the ventral pathway but this is only possible for words and not nonwords).

The Nature of Computed Representations Depends on Neural Location and Task

Analysis of each layer in the model demonstrated that the internal similarity structure changed gradually across successive regions. In line with recent neuroimaging results (Scott et al., 2000; Visser and Lambon Ralph, 2011), the ventral pathway shifted from coding predominantly acoustic/phonological to predominantly semantic structure. Additional control simulations (comparing this multilayer pathway with a single, larger intermediate layer; see Figure S3) indicated that this gradual shift led to much better performance when extracting the modality-invariant meaning from the time-varying auditory input. Finally, a second key finding from these analyses is that the structure of the representations can change across tasks even within the same region. For example, the aSTG is much more sensitive to semantic similarity during speaking/naming than in comprehension, a fact that might explain recent VSLM data (Schwartz et al., 2009) (see Results). If correct, then this result has clear implications for the limits of the subtraction assumption (Price et al., 1997), commonly utilized in functional neuroimaging.

EXPERIMENTAL PROCEDURES

Rationale for Methodological Details (Working Assumptions)

When implementing any cognitive or neural hypothesis in a computational model various assumptions have to be made explicit. In this section we outline our working assumptions and the rationale underlying them. We then provide a summary of implementational details. Copies of the model files are available from the authors upon request.

Level of Granularity

The model was aimed at explaining language performance and the behavioral neurology of aphasic patients in terms of the interaction between widely-distributed brain systems. Accordingly, the model was implemented at the systems level (i.e., each layer of units is expected to reflect the functioning of a specific brain region) rather than at the micro level of neuronal assemblies or spiking neurons.

Layers and Connections

We assume a layer represents a cortical region which computes representations and delivers information through its ongoing connections (O'Reilly, 2006). Connections primarily represent white matter pathways and language processing is underpinned by both cortical regions and their connectivity (Mesulam, 1990). Like real cortical areas, the layers have both afferent and efferent connections.

Nonspecified Representations

Other than the representations applied at the input or output layers, the rest of the model's function was unspecified. In this sense, these representations are not present at onset but are formed across the intermediate units and connections in order to maximize performance across the various tasks. Following development or recovery, the nature of the resultant representations has to be probed by the modeler.

Prespecified Representations

Three layers of the model were assumed to be the starting (input) and end (output) points of the simulated language activities and so the representations for these regions were prespecified. The primary auditory area and surrounding region, including pSTG, process complex acoustic stimuli including phonetic contrasts (Chang et al., 2010; Griffiths, 2002). Accordingly, the corresponding input layer of the model coded phonetic-based auditory inputs for all the words in the training set and novel forms (for testing generalization). Anterior insular cortex has been demonstrated to play a key role in speech output (Baldo et al., 2011; Dronkers, 1996; Wise et al., 1999). Although classically implicated in speech, the role of pars opercularis is more controversial (Dhanjal et al., 2008; Wise et al., 1999). As a result, we assume that this general insular-motor area plays a key role in speech output and so the corresponding layer in the model was set to generate speech output. Finally, inferolateral (ventral) anterior temporal cortex (vATL) is known to be a key locus for transmodal semantic representations and thus crucial for both multimodal comprehension and the semantic input to speech production/naming (Lambon Ralph et al., 2001; Rogers et al., 2004; Visser and Lambon Ralph, 2011). This is not to say that this is the only key region for semantic cognition. Indeed, other regions provide modality-specific sources of information or executive mechanisms for controlled semantic processing (Jefferies and Lambon Ralph, 2006). Unlike more complex tasks or nonverbal semantic processing, these components of semantic cognition are not crucial to the single-word comprehension and speaking/naming tasks included in the model's training regime. Thus, in order to keep an already complex computational framework manageable, we implemented the vATL semantic system alone. Specifically, it was set to generate semantic outputs for comprehension and provided the semantic input for speaking/naming. In repetition, this layer was not assigned a specific role and so its activations were unconstrained.

Core Representational Characteristics

The prespecified representations were designed to capture some of the most computationally-demanding and fundamental characteristics of processing in each domain. One of the major challenges in auditory processing and speech production is to deal with time-varying inputs and outputs. In repetition, for example, the sequentially-incoming auditory input has to be amalgamated and then used for reproduction in the correct order (Plaut and Kello, 1999). Another key characteristic is that at any one point of the auditory stream, there are multiple phonetic features to be processed (e.g., fricative, sonorant, etc.) (Plaut and Kello, 1999). Our representations conformed to these two demands by coding the acoustic-phonological input and phonetic-motor output as time-varying, phonetic-based distributed representations (see Supplemental Experimental Procedures, for the details of the coding methodology). Identical vectors were used for speech input and output, even though there probably should be acoustic-/articulatory-specific factors (Plaut and Kello, 1999). In order to keep the complex simulation manageable, however, we skipped acoustic-analysis and articulation phases.

In contrast, conceptual knowledge is both time- and modality-invariant (Lambon Ralph et al., 2010; Rogers et al., 2004) and our semantic representations conformed to these two demanding computational requirements. Specifically, the network was pressured to compute the time-invariant semantic information as soon as possible after the onset of the auditory input (Plaut and Kello, 1999). Likewise for speech production, the same time-invariant semantic representation was used to generate time-varying, distributed phonetic output. In addition, the mapping between auditory input/speech output and semantic representations is arbitrary in nature and this provides an additional challenge to any computational model (Rogers et al., 2004). Accordingly, we ensured that the similarity structure of the semantic representations was independent of the auditory input/speech output representations.

Unlike speech, which is an external stimulus and present in the environmental throughout a human's lifespan, semantic knowledge is internally represented and gradually accumulated during development. Accordingly, like past computational models, the current study assumed that (1), children gradually develop their internal semantic representations (Rogers et al., 2004) and (2), at any time point of their development, children use the current, "developing"

internal semantic representations to drive spontaneous speaking (Plaut and Kello, 1999; see Supplemental Experimental Procedures).

Language-Related Areas and Connectivity

Figure 1 shows the structure of the implemented dual-pathway language model, which was constrained by the following recent neuroscience findings. As was noted in the main text, Wernicke's area and Broca's area are connected via both the AF (Geschwind, 1970; Parker et al., 2005; Saur et al., 2008) and EmC (Parker et al., 2005; Saur et al., 2008). Recent in vivo MR tractography has added further information about each pathway. Specifically, AF divides into several branches, of which the most language-related one starts from primary auditory and pSTG, and projects to the insular cortex as well as Broca's area (Bernal and Ardila, 2009; M.A.L.R. et al., unpublished data). This AF branch passes through and connects to the inferior supramarginal gyrus (iSMGSMG) (Parker et al., 2005; M.A.L.R. et al., unpublished data), which plays a critical role in human phonological processing (Hartwigsen et al., 2010) and acts as a sound-motor interface in primates (Rauschecker and Scott, 2009).

The ventral pathway is underpinned initially by the middle longitudinal fasciculus (MLF), connecting primary auditory and pSTG to mSTG and aSTG. At this point there is a bifurcation, with the EmC branch connecting to inferior prefrontal regions (pars triangularis and opercularis; M.A.L.R. et al., unpublished data; Parker et al. [2005]). The vATL is not directly connected to the prefrontal cortex but is strongly connected to other temporal lobe regions including the aSTG (M.A.L.R. et al., unpublished data). In addition to the EmC, anterior temporal, and especially temporal polar regions, are connected to the pars orbitalis and orbitofrontal areas via the UF (M.A.L.R. et al., unpublished data). While it is possible that this connection may play a role in language or semantic function, direct stimulation studies indicates that the EmC is crucial for spoken language (Duffau et al., 2009) and thus this connection was implemented. Finally, we split the STG layer into two in the model in order to capture the functional transition along the rostral STG/STS (Scott et al., 2000) (see Aims). In reality, this shift is likely to be much more gradual in form but, for the sake of computational simplicity, we split the layer into two parts.

Tasks

We focused on the major language activities of single-word repetition, comprehension and speaking/naming, which play a key role in differential diagnosis of the principal aphasia types. Multiple-word processing (e.g., connected speech and serial order recall) is a future target. Although we did not train the model to repeat nonwords, it was tested on these novel items in order to assess the model's generalization of acoustic-motor statistical information.

Lesion Simulation

Almost all forms of brain damage involve both cortical regions and underlying white matter—indicating that most neuropsychological disorders reflect a combination of cortical dysfunction and disconnection. Therefore, damage was simulated by removing a proportional of the incoming links to the damaged layer (reducing its connectivity) and also by applying noise over the output of that layer (damaging the functioning of the "cortical" region). The varying severity of the patients' impairments was simulated by altering the degree of damage.

Summary of Implementation Details

Architecture and the Activation Flow

A neural network model was constructed and trained with the Light Efficient Network Simulator (Rohde, 1999). It was implemented as a simple-recurrent Elman network model to reduce the computational demands (Plaut and Kello, 1999) and the exact computational architecture to realize this implementation is shown in Figure S1. Specifically, once a pattern was clamped to the sound input layer (in repetition/comprehension, for example), the activation spread to (1), iSMG → insular-motor layers; (2), to mSTG → aSTG → vATL layers; and (3), mSTG → aSTG → triangularis-opercularis → insular-motor layers at every time tick. The activation pattern at every layer was fed back to the previous layer at the next time tick by utilizing the copy-back connections to realize bidirectional connectivity (see Figure S1 for further details). A sigmoid activation function was used for each unit, bounding activation at 0 and 1.

Training Patterns

Eight hundred fifty-five high-frequency and eight hundred fifty-five low-frequency Japanese nouns, each three moras (subsyllabic spoken unit) in length, were selected from the NTT Japanese psycholinguistic database (Amano and Kondo, 1999)(see Supplemental Experimental Procedures, for the item properties). The remaining 3511 tri-mora nouns in the corpus were used for testing generalization. Each mora was converted into a vector of 21 bits representing pitch accent and the distinctive phonetic-features of its comprising consonant and vowel (the exact vector patterns are provided in Supplemental Experimental Procedures), following previous coding systems (Halle and Clements, 1983). Past simulations of language activities in English have used exactly the same coding scheme (Harm and Seidenberg, 2004) and so our findings should be language-general. The acoustic/motor representation of each word was made up of the three sequential, distributed mora patterns. Semantic representations were abstract vector patterns of 50 bits generated from 40 prototype patterns, containing 20 "on" bits randomly dispersed across the 50 semantic units. Fifty exemplar patterns were generated from each of these prototypes by randomly turning off 10 of the 20 on bits, again following the previous coding systems of past English simulations (Plaut et al., 1996). Each semantic pattern was randomly assigned to one of the 1710 auditory patterns, ensuring an arbitrary mapping between the two types of representation.

Training and Testing

In repetition, each 21-bit mora vector was clamped to the input auditory layer, sequentially (i.e., one mora per tick), during which the insular-motor output layer was required to be silent. From the fourth tick (once the entire word had been presented), the output layer was trained to generate the same vector patterns sequentially (i.e., one mora per a tick), resulting in six time ticks in total. The simulation was required, therefore, to "hear" all of the word before beginning to repeat it, thus requiring some degree of phonological buffering. In comprehension, the three 21-bit mora input vectors for each word were clamped in the same way (i.e., three ticks), during which the target semantic pattern was compared to the output of the vATL layer at every time tick (i.e., a time-varying to time-invariant transformation). During comprehension trials, the insular-motor speech output layer was required to be silent. In speaking/naming, the developing semantic pattern was clamped to the vATL layer for three time ticks, during which the insular-motor output layer generated the three 21-bit mora vectors sequentially (i.e., a time-invariant to time-varying transformation).

During every epoch of training, each word appeared once for repetition (1/6), two times for speaking (2/6), and three times for comprehension (3/6) in a random order. Note that the order of acquisition observed in the model is not attributable to these frequency choices, as the model learned the less frequent production task (repetition) prior to the more frequent production task (naming). The network updated the connection weights after every item (i.e., online learning) using the standard backpropagation algorithm. Performance was evaluated after every 20 epochs, where an output was scored as correct when the activation in every unit of the output layer was in the correct side of 0.5 (i.e., on units should be >0.5, whereas "off" units should be <0.5). Comprehension accuracy was evaluated on the output in the last tick, at which point the network had received all of the three 21-bit mora input vectors (i.e., the whole word). Training finished at epoch 200, at which point 2.05 million words had been presented. It is difficult to know exactly how to scale and map between training time/epochs in a model to developmental time in children. Plaut and Kello (1999) noted that they trained their model of spoken language processing on 3.5 million word presentations. They argued "although this may seem like an expressive amount of training, children speak up to 14,000 words per day (Wagner, 1985, *Journal of Child Language*), or over 5 million words per year." Our training length (~2 million word presentations) is far less than this.

Other Training Parameters

Random Seeds. Five networks were trained independently with different random seeds (different initial weight values). The data reported in the figures/tables is the average of the results over these five independent simulations (and standard errors), except for Figure 6 where ten simulations were used.

Learning Rate and Weight Decay. The training was initiated with a learning rate of 0.5 until the end of epoch 150. After this, the learning rate was gradually reduced by 0.1 per 10 epochs to the end of epoch 180 (at this point, the learning rate was fixed at 0.1). Training finished at epoch 200. Weight decay was adjusted using the same schedule. It was initiated at 10^{-6} and reduced

by 10^{-7} per 10 epochs from the end of epoch 150 to the end of epoch 180, and then kept at 6×10^{-7} until epoch 200.

Word Frequency. Log_{10} transformed frequency of each word was used to scale the error derivatives. This level of frequency compression was employed to reduce the training time in this large model (Plaut et al., 1996). The error (difference between the target and the output patterns) was estimated with the cross-entropy method (Hinton, 1989).

Zero-Error Radius. No error was backpropagated from a unit if the difference between the output and the target was <0.1 (i.e., zero-error radius parameter was set to 0.1).

Momentum and Initial Weights. Momentum was not used in this study. All the units in the hidden and output layers had a trainable bias link, except for the copy and Elman layers. Weights were initialized to random values between −1 and 1 (0.5 for recurrent connections). Weights from the bias unit to hidden units were initialized at −1, so as to avoid strong hidden unit activation early in training (Botvinick and Plaut, 2006). A sigmoid activation function was used for each unit with activation ranging from 0 to 1. At the beginning of each trial, activations for all units in the hidden layer (including vATL-output layer) were set to 0.5, and for all units in the insular-motor output layer to zero.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and three figures and can be found with this article online at doi:10.1016/j.neuron.2011.09.013.

ACKNOWLEDGMENTS

This work was supported by an MRC programme grant to M.A.L.R. (G0501632), a Royal Society travel grant to M.A.L.R. and S.S., and a Study Visit grant from the Experimental Psychology Society to T.U. T.U. was supported by an Overseas Research Scholarship (UK) and an Overseas Study Fellowship from the Nakajima Foundation (Japan).

Accepted: September 14, 2011
Published: October 19, 2011

REFERENCES

Amano, S., and Kondo, T. (1999). Japanese NTT database series: lexical properties of Japanese (1) (Tokyo: Sanseido).

Baldo, J.V., Wilkins, D.P., Ogar, J., Willock, S., and Dronkers, N.F. (2011). Role of the precentral gyrus of the insula in complex articulation. Cortex *47*, 800–807.

Bernal, B., and Ardila, A. (2009). The role of the arcuate fasciculus in conduction aphasia. Brain *132*, 2309–2316.

Binder, J.R., Gross, W.L., Allendorfer, J.B., Bonilha, L., Chapin, J., Edwards, J.C., Grabowski, T.J., Langfitt, J.T., Loring, D.W., Lowe, M.J., et al. (2011). Mapping anterior temporal lobe language areas with fMRI: a multicenter normative study. Neuroimage *54*, 1465–1475.

Binney, R.J., Embleton, K.V., Jefferies, E., Parker, G.J.M., and Ralph, M.A. (2010). The ventral and inferolateral aspects of the anterior temporal lobe are crucial in semantic memory: evidence from a novel direct comparison of distortion-corrected fMRI, rTMS, and semantic dementia. Cereb. Cortex *20*, 2728–2738.

Botvinick, M.M., and Plaut, D.C. (2006). Short-term memory for serial order: a recurrent neural network model. Psychol. Rev. *113*, 201–233.

Chang, E.F., Rieger, J.W., Johnson, K., Berger, M.S., Barbaro, N.M., and Knight, R.T. (2010). Categorical speech representation in human superior temporal gyrus. Nat. Neurosci. *13*, 1428–1432.

Crisp, J., and Lambon Ralph, M.A. (2006). Unlocking the nature of the phonological-deep dyslexia continuum: the keys to reading aloud are in phonology and semantics. J. Cogn. Neurosci. *18*, 348–362.

Dell, G.S., Schwartz, M.F., Martin, N., Saffran, E.M., and Gagnon, D.A. (1997). Lexical access in aphasic and nonaphasic speakers. Psychol. Rev. *104*, 801–838.

Dhanjal, N.S., Handunnetthi, L., Patel, M.C., and Wise, R.J.S. (2008). Perceptual systems controlling speech production. J. Neurosci. *28*, 9969–9975.

Dilkina, K., McClelland, J.L., and Plaut, D.C. (2008). A single-system account of semantic and lexical deficits in five semantic dementia patients. Cogn. Neuropsychol. *25*, 136–164.

Dilkina, K., McClelland, J.L., and Plaut, D.C. (2010). Are there mental lexicons? The role of semantics in lexical decision. Brain Res. *1365*, 66–81.

Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. Nature *384*, 159–161.

Duffau, H., Gatignol, P., Moritz-Gasser, S., and Mandonnet, E. (2009). Is the left uncinate fasciculus essential for language? A cerebral stimulation study. J. Neurol. *256*, 382–389.

Eggert, G.H. (1977). Wernicke's Works on Aphasia: A Sourcebook and Review (The Hague: Mouton).

Fridriksson, J., Kjartansson, O., Morgan, P.S., Hjaltason, H., Magnusdottir, S., Bonilha, L., and Rorden, C. (2010). Impaired speech repetition and left parietal lobe damage. J. Neurosci. *30*, 11057–11061.

Galton, C.J., Patterson, K., Graham, K., Lambon-Ralph, M.A., Williams, G., Antoun, N., Sahakian, B.J., and Hodges, J.R. (2001). Differing patterns of temporal atrophy in Alzheimer's disease and semantic dementia. Neurology *57*, 216–225.

Geschwind, N. (1965). Disconnexion syndromes in animals and man. I. Brain *88*, 237–294.

Geschwind, N. (1970). The organization of language and the brain. Science *170*, 940–944.

Griffiths, T.D. (2002). Central auditory processing disorders. Curr. Opin. Neurol. *15*, 31–33.

Halle, M., and Clements, G. (1983). Problem Book in Phonology (Cambridge, MA: MIT Press).

Harm, M.W., and Seidenberg, M.S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. Psychol. Rev. *111*, 662–720.

Hartwigsen, G., Baumgaertner, A., Price, C.J., Koehnke, M., Ulmer, S., and Siebner, H.R. (2010). Phonological decisions require both the left and right supramarginal gyri. Proc. Natl. Acad. Sci. USA *107*, 16494–16499.

Hickok, G., and Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. (Regul. Ed.) *4*, 131–138.

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. Nat. Rev. Neurosci. *8*, 393–402.

Hillis, A.E. (2007). Aphasia: progress in the last quarter of a century. Neurology *69*, 200–213.

Hinton, G.E. (1989). Connectionist learning procedures. Artif. Intell. *40*, 185–234.

Hodges, J.R., Patterson, K., Oxbury, S., and Funnell, E. (1992). Semantic dementia. Progressive fluent aphasia with temporal lobe atrophy. Brain *115*, 1783–1806.

Holland, R., and Lambon Ralph, M.A. (2010). The anterior temporal lobe semantic hub is a part of the language neural network: selective disruption of irregular past tense verbs by rTMS. Cereb. Cortex *20*, 2771–2775.

Jefferies, E., and Lambon Ralph, M.A. (2006). Semantic impairment in stroke aphasia versus semantic dementia: a case-series comparison. Brain *129*, 2132–2147.

Jefferies, E., Patterson, K., Jones, R.W., and Lambon Ralph, M.A. (2009). Comprehension of concrete and abstract words in semantic dementia. Neuropsychology *23*, 492–499.

Lambon Ralph, M.A. (2010). Measuring language recovery in the underlying large-scale neural network: Pulling together in the face of adversity. Ann. Neurol. *68*, 570–572.

Lambon Ralph, M.A., Graham, K.S., Ellis, A.W., and Hodges, J.R. (1998). Naming in semantic dementia—what matters? Neuropsychologia *36*, 775–784.

Lambon Ralph, M.A., McClelland, J.L., Patterson, K., Galton, C.J., and Hodges, J.R. (2001). No right to speak? The relationship between object naming and semantic impairment: neuropsychological evidence and a computational model. J. Cogn. Neurosci. *13*, 341–356.

Lambon Ralph, M.A., Sage, K., Jones, R.W., and Mayberry, E.J. (2010). Coherent concepts are computed in the anterior temporal lobes. Proc. Natl. Acad. Sci. USA *107*, 2717–2722.

Leff, A., Crinion, J., Scott, S., Turkheimer, F., Howard, D., and Wise, R. (2002). A physiological change in the homotopic cortex following left posterior temporal lobe infarction. Ann. Neurol. *51*, 553–558.

Lichtheim, L. (1885). On aphasia. Brain *7*, 433–485.

Matsumoto, R., Nair, D.R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., and Lüders, H.O. (2004). Functional connectivity in the human language system: a cortico-cortical evoked potential study. Brain *127*, 2316–2330.

McClelland, J.L., St. John, M., and Taraban, R. (1989). Sentence comprehension: a parallel distributed processing approach. Lang. Cogn. Process. *4*, SI287–SI335.

Mesulam, M.M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. Ann. Neurol. *28*, 597–613.

Mion, M., Patterson, K., Acosta-Cabronero, J., Pengas, G., Izquierdo-Garcia, D., Hong, Y.T., Fryer, T.D., Williams, G.B., Hodges, J.R., and Nestor, P.J. (2010). What the left and right anterior fusiform gyri tell us about semantic memory. Brain *133*, 3256–3268.

Nozari, N., Kittredge, A.K., Dell, G.S., and Schwartz, M.F. (2010). Naming and repetition in aphasia: Steps, routes, and frequency effects. J. Mem. Lang. *63*, 541–559.

O'Reilly, R.C. (2006). Biologically based computational models of high-level cognition. Science *314*, 91–94.

Parker, G.J.M., Luzzi, S., Alexander, D.C., Wheeler-Kingshott, C.A.M., Ciccarelli, O., and Lambon Ralph, M.A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. Neuroimage *24*, 656–666.

Patterson, K., Nestor, P.J., and Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. Nat. Rev. Neurosci. *8*, 976–987.

Plaut, D.C., and Kello, C.T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In Emergence of Language, B. MacWhinney, ed. (Mahwah: Lawrence Erlbaum), pp. 381–415.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., and Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. Psychol. Rev. *103*, 56–115.

Pobric, G., Jefferies, E., and Ralph, M.A. (2007). Anterior temporal lobes mediate semantic representation: mimicking semantic dementia by using rTMS in normal participants. Proc. Natl. Acad. Sci. USA *104*, 20137–20141.

Price, C.J., Moore, C.J., and Friston, K.J. (1997). Subtractions, conjunctions, and interactions in experimental design of activation studies. Hum. Brain Mapp. *5*, 264–272.

Rauschecker, J.P., and Scott, S.K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Neurosci. *12*, 718–724.

Rogers, T.T., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R., and Patterson, K. (2004). Structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychol. Rev. *111*, 205–235.

Rohde, D.L.T. (1999). LENS: the light, efficient network simulator. Technical Report CMU-CS-99-164 (Pittsburgh, PA: Carnegie Mellon University, Department of Computer Science).

Saur, D., Kreher, B.W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.S., Umarova, R., Musso, M., Glauche, V., Abel, S., et al. (2008). Ventral and dorsal pathways for language. Proc. Natl. Acad. Sci. USA *105*, 18035–18040.

Schwartz, M.F., Kimberg, D.Y., Walker, G.M., Faseyitan, O., Brecher, A., Dell, G.S., and Coslett, H.B. (2009). Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. Brain *132*, 3411–3427.

Scott, S.K., Blank, C.C., Rosen, S., and Wise, R.J.S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. Brain *123*, 2400–2406.

Seidenberg, M.S., and McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. Psychol. Rev. *96*, 523–568.

Shallice, T., and Warrington, E.K. (1977). Auditory-verbal short-term memory impairment and conduction aphasia. Brain Lang. *4*, 479–491.

Sharp, D.J., Scott, S.K., and Wise, R.J.S. (2004). Retrieving meaning after temporal lobe infarction: the role of the basal language area. Ann. Neurol. *56*, 836–846.

Sharp, D.J., Turkheimer, F.E., Bose, S.K., Scott, S.K., and Wise, R.J.S. (2010). Increased frontoparietal integration after stroke and cognitive recovery. Ann. Neurol. *68*, 753–756.

Visser, M., and Lambon Ralph, M.A. (2011). Differential contributions of bilateral ventral anterior temporal lobe and left anterior superior temporal gyrus to semantic processes. J. Cogn. Neurosci. *23*, 3121–3131.

Weiller, C., Bormann, T., Saur, D., Musso, M., and Rijntjes, M. (2011). How the ventral pathway got lost: and what its recovery might mean. Brain Lang. *118*, 29–39.

Welbourne, S.R., and Lambon Ralph, M.A. (2007). Using parallel distributed processing models to simulate phonological dyslexia: the key role of plasticity-related recovery. J. Cogn. Neurosci. *19*, 1125–1139.

Wise, R.J.S., Greene, J., Büchel, C., and Scott, S.K. (1999). Brain regions involved in articulation. Lancet *353*, 1057–1061.

Woollams, A.M., Ralph, M.A., Plaut, D.C., and Patterson, K. (2007). SD-squared: on the association between semantic dementia and surface dyslexia. Psychol. Rev. *114*, 316–339.

Neuron, Volume *72*

Supplemental Information

Lichtheim 2: Synthesizing Aphasia and the Neural Basis of Language in a Neurocomputational Model of the Dual Dorsal-Ventral Language Pathways

Taiji Ueno, Satoru Saito, Timothy T. Rogers, and Matthew A. Lambon Ralph

Inventory

These supplementary materials support two important aspects of the main paper. Computational simulations can be considered as a series of 'computational experiments'. Like any other laboratory-based experiments, it is sometimes necessary to conduct 'control' experiments in order to check various assumptions or methods. Accordingly, the first part (Supplementary Data) provides these 'control simulations'. Secondly, like other complex laboratory experiments, a computational model involves a multitude of set-up parameters and other detail. Therefore, the Supplemental Experimental Procedures provide all of these details such that any other computational modeller could re-implement the model for replication or extension of the 'computational experiments' in the future. Both sections are important to validate our main findings but not central to be included into the main text.

Supplementary Data

Supplementary Experimental Procedures

- This describes the nature of Japanese phonology, item selection and how phonological and semantic representations were created.

(5) *Lesioning procedure (cf. Figure 3)* … *(pp.14-16)*

- This provides additional details of the experimental procedure used for lesioning and synthesis of aphasia types (data in Figure 3).

(6) *Simulation of recovery (cf. Figure 4)* … *(p.17)*

- This sets out additional information for the training and analysis of the recovery simulation (Figure 4).

(7) *Analysis of the encoded similarity structure within each computational layer (cf. Figure 5).* … *(pp.18-19)*

- This describes the method for investigating the similarity structure across each computational layer (Figure 5).

(8) *Semantic naming errors (cf. Figure 6)* … *(pp.20-21)*

- This provides details of how we probed for semantic errors in the model after different locations of damage (Figure 6).

Supplementary References … *(p.22)*

Supplemental Data

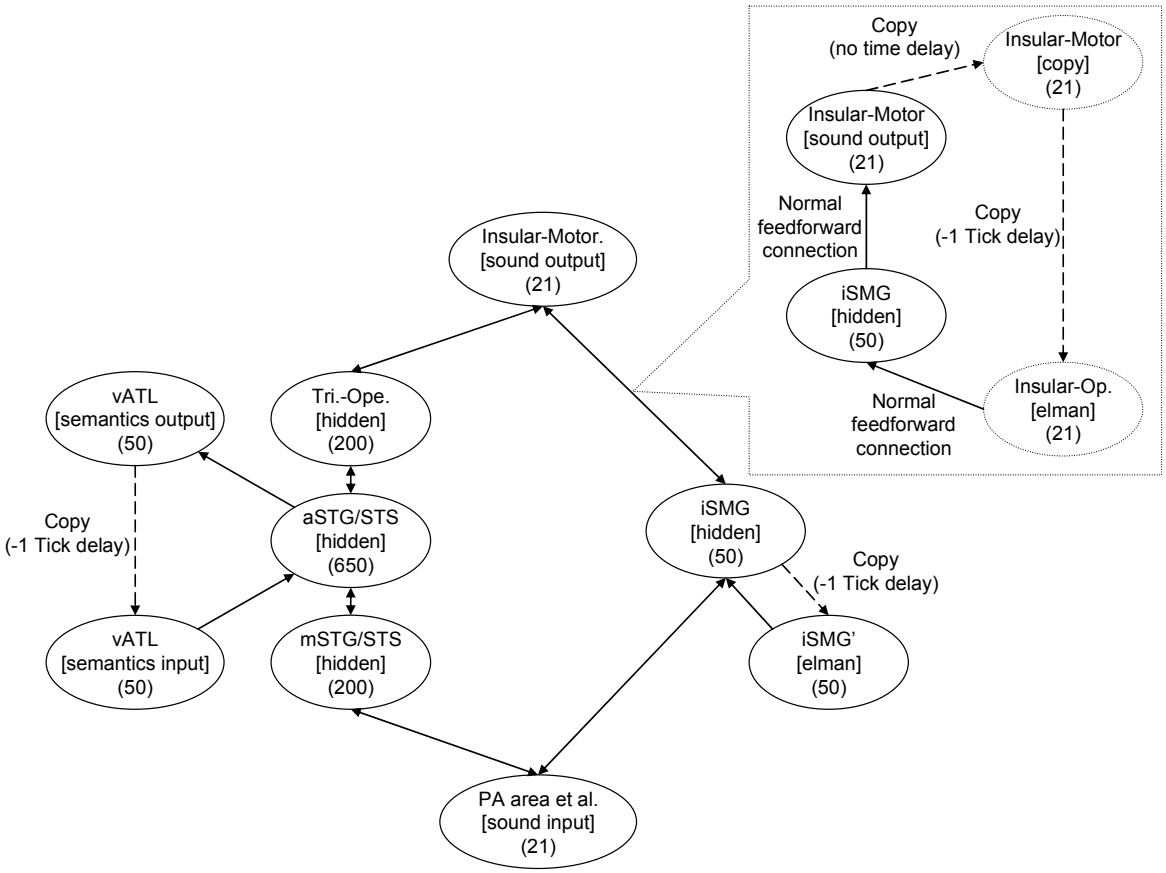

Figure S1 (related to Figure 1). Implemented computational model and the activation flow.

The network was built up and trained with the software LENS (http://tedlab.mit.edu/~dr/Lens/) (Rohde, 1999). Figure S1 shows the exact computational architecture used to realize the bidirectional connections of the main simulation (related to Figure 1, Main Text) by utilising a simple-recurrent Elman network model (directly following the "unfolding" method as developed by Plaut and his collaborators (Botvinick and Plaut, 2006; Plaut, 1997; Plaut and Kello, 1999)). Arrows in the upward direction of Figure S1 represent the feedforward connections. Thus, for example, once a pattern was clamped to the sound input layer in repetition, the activation spread to the other layers and every layer was updated at every time tick.

Feedback connections (repetition).

In the next time tick, when a new input was clamped to the sound-input layer, activation was spread in the same way. Simultaneously, feedback was implemented by taking the activation patterns on the various layers from the previous time tick (copied to their invisible context layers) and sending these activations back to the previous layers in the

network. Specific details on the form of the feedback in the various parts of the model are provided below.

(1) *Recurrent self-connection (Elman connection in the iSMG layer)*: The activation pattern on the iSMG layer in the 1st tick was copied to its Elman context layer (see Figure S1), and in the 2nd tick was sent back to the iSMG layer. This combination allows the iSMG circuit to represent both the new input and its own activation pattern from the previous time point, thus realizing a memory buffering function(Elman et al., 1996). This operated on every tick of every task.

(2) *Copy from the vATL-output layer to the vATL-input layer*. The vATL layer was divided into two in order to realize the feedback connection to aSTG. At every time tick, the output pattern from the vATL-output layer was copied and hard-clamped to the vATL-input layer on the next tick, and this activation propagated to the aSTG layer simultaneously with the activation arriving from the input-sound layer.

(3) *Other feedback connections (invisible context layers).* Realization of the other "feedback" connections, using the Plaut unfolding method, is exemplified in the upper-right box of Figure S1. For example, the 1st tick activation pattern in the insular-motor layer was duplicated on its copy layer and this was used to feedback onto the iSMG layer in the 2nd tick. This operated at every time tick of every task. This method was used for all of the bidirectional arrows shown in Figure S1.

Comprehension

In comprehension, the input sound pattern was hard-clamped in the same way as for repetition and the target pattern was applied to the vATL-output layer during the three time ticks. The copy and Elman connections operated in the same manner as in repetition.

Speaking

In speaking, the developing semantic pattern (see below) was hard-clamped to the vATL-input layer and the three mora target pattern applied to the insular-motor output layer over three time ticks. Since this input pattern was hard-clamped, the copy from the vATL-output layer to vATL-input layer was ignored. The operation of the other copy and Elman connections was identical to that in comprehension and in repetition.

Developing semantic representations.

As mentioned in the main text, the developing semantic representations were used as the input pattern to speaking. Specifically, after every comprehension trial, the activation at the vATL-output layer for each word (measured at the 3rd tick) was updated and saved (as 'objects' in LENS, http://tedlab.mit.edu/~dr/Lens/objects.html). Then, for the next speaking trial, this saved developing semantic pattern was hard-clamped to the vATL-input layer.

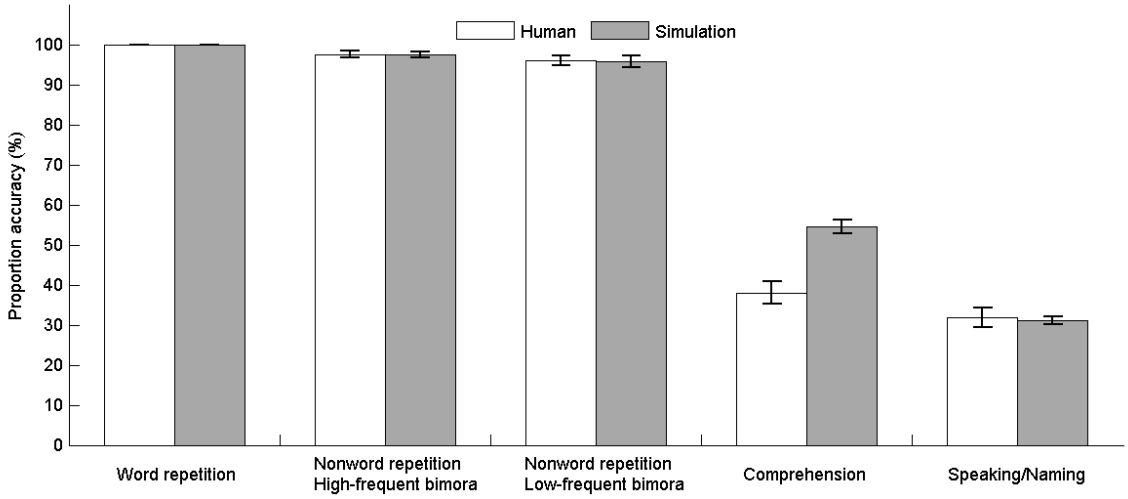

Figure S2 (related to Figure 2). Comparisons with human normative data across five language tasks. Human nonword repetition was derived from Tanida et al. (in preparation)

Figure S2 compares the results of human behavioural experiments and those of the simulation, in which identical items were used. In every task, the network achieved equivalent performance to humans (or even better than humans).

Word Repetition.

We assessed repetition performance simultaneously with the auditory-word comprehension normative experiment (see below). The human and network performance was perfect.

Nonword repetition.

Tanida et al. (in preparation) created 36 high-bimora-frequent and 36 low-bimora frequent Japanese nonwords (bi-mora frequency was taken from a Japanese database (Tamaoka and Makioka, 2004). Each of 72 tri-mora nonwords was presented three times (once in a block), in each of which different pitch accent patterns were assigned (i.e., 216 nonword trials: 72 mora sequences × 3 accent types). The participants were required to repeat the nonword with the same accent pattern without time pressure. The network was presented with exactly the same 216 nonword items. Repetition performance was averaged across the three accent types. In both high- and low-bimora-frequency conditions, the network and human performances were very similar. In Figure 2 (main text), the learning curve for nonword repetition reflects the accuracy in the high-bimora-frequent condition.

Comprehension.

Figure 2 (main text) shows that comprehension and speaking in the network were not perfect for low-frequent words. Thus, in order to compare with the human comprehension

performance, we extracted 120 lower word-frequent items with two constraints: Firstly, words were collected from three accent type categories as evenly as possible while prioritizing the items with lower word-frequency (60 Flat words, 40 Type1-words, and 20 Type-2 words; Mean $\log_{10}$-transformed word-frequency = 0.796, *SD* = 0.470). Secondly, ethically-unfavourable words were removed.

Six students listened to an auditory word, one at a time. To make sure if they listened to it correctly, they were at first required to repeat it with the presented pitch accent (thus, we obtained the repetition performance simultaneously, see above). Then, the participants were required to define the meaning of each word. The response was scored correct if they could produce at least one correct attribute. In spite of this generous scoring criterion, the human performance was lower than the network (in most trials, they simply said “do not know”, which was not surprising given the very low word-frequency of the items used).

Speaking/naming.

The same 120 low frequency items were used. The meaning of each word was shown in a PowerPoint slide. Nine students were shown each slide and were asked to provide a tri-mora name that matched the presented meaning. When they could not generate a correct name, the initial mora was cued. If the participant could generate with this cue, the response was scored correct. With this generous criterion, human performance was just comparable to the network.

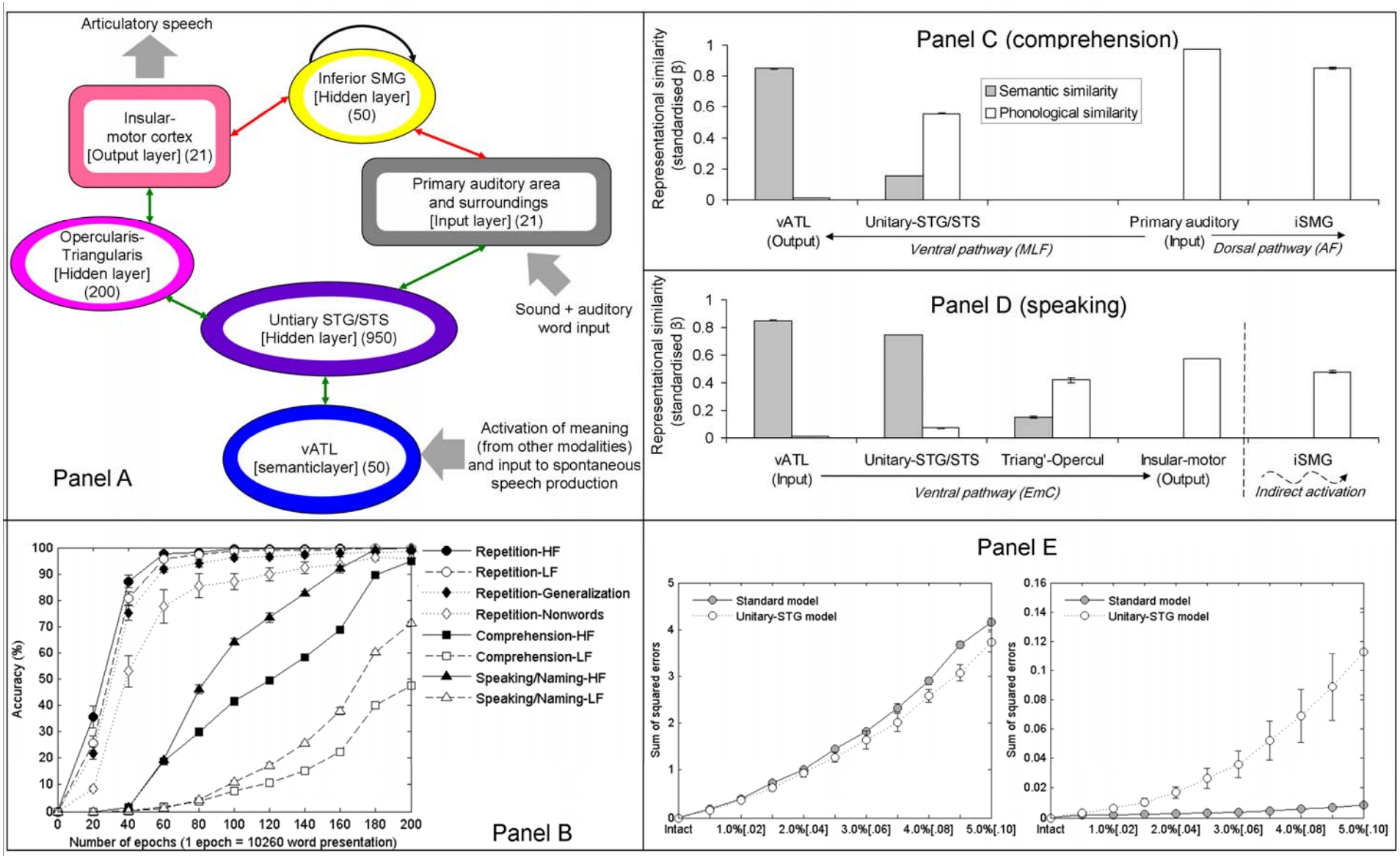

Figure S3 (related to Figure 7). Architecture of a model with a unitary STG layer. Panel A - accuracy of three language tasks during training. Panel B - analysis of representational similarity during comprehension (Panel C) and speaking (Panel D) within the model. The effects of diagnostic lesioning on repetition (Panel E) in the iSMG layer (left panel) or in the vATL layer (right panel) of the two models. See the following text for explanations of each panel.

Architecture (Panel A)

The purpose of this additional simulation was to explore the impact of including two sequential layers in the STG. Accordingly a new model with a single STG layer was trained and tested for comparison (see *Panel A of Figure S3,* above). The connections and training parameters were implemented in the same way. The number of processing units in the unitary-STG layer was set to 950 so that total number of links was matched between unitary-STG network (614,092) and the standard network (610,792).

Learning curves (Panel B).

Panel B of Figure S3 shows the learning curves for the unitary-STG model. Compared to the standard model, this simulation exhibited poorer comprehension and speaking/naming performances for both high and low frequent words. The underlying change in the nature of the representations in this model was explored using the representational similarity analyses, as used to probe the standard model.

Regression planes (Panels C & D)

Panels C and D of Figure S3 summarise the results of the multiple-regression analysis, conducted for unitary-STG model. Our focus here is on the Unitary-STG layer. This layer captured less semantic similarity (β= .15) than the aSTG layer (β= .23) and also less phonological similarity (β= .55) than the mSTG layer (β= .59) in the standard model. Thus, these analyses suggest that the unitary-STG model "compressed" both the phonological processing and semantic processing within one layer, resulting in worse comprehension and naming performance. This finding suggests that it may be critical or at least more efficient to shift from the time-variant acoustic-phonological to time-invariant semantic representation in a gradual manner as would appear to be the case in the human STG/STS (Plaut, 2002; Scott et al., 2000).

Residual errors (Panel E)

An additional lesion-based analysis suggests that there is also a changed division of labour (between the two pathways) when the ventral route only contains one STG layer. Panel E of Figure S3 shows the effect on repetition of a diagnostic lesioning in the iSMG layer (left) or in the vATL layer (right) of the two models.

The Y-axis represents the amount of remaining sum-of-squared errors (the difference between the target and output). Higher values indicate poorer performance. Note here that the y-axis scale is different for the two figures in this panel because, in both models, repetition was mainly supported by the dorsal pathway. Critically, dorsal lesioning had a bigger effect in the standard model than the unitary-STG model whereas the opposite was true in case of vATL damage. This means that repetition was more reliant on the ventral pathway in the unitary-STG model, presumably because the unitary layer was primarily

capturing acoustic-phonological similarity (see above), at the cost of somewhat lower comprehension accuracy.

Supplemental Experimental Procedures

Japanese phonology, item selection and representations

Table: All Japanese mora (within the thick rectangle)

		Vowels				
		[a]	[e]	[i]	[o]	[ɯ]
Consonants	##	[a]	[e]	[i]	[o]	[ɯ]
	[k]	[ka]	[ke]		[ko]	[kɯ]
	[kʲ]	[kʲ]		[kʲi]	[kʲo]	[kʲɯ]
	[g]	[ga]	[ge]		[go]	[gɯ]
	[gʲ]	[gʲa]		[gʲi]	[gʲo]	[gʲɯ]
	[s]	[sa]	[se]		[so]	[sɯ]
	[ʃ]	[ʃa]		[ʃi]	[ʃo]	[ʃɯ]
	[ʑ]	[ʑa]	[ʑe]		[ʑo]	[ʑɯ]
	[ɕ]	[ɕa]		[ɕi]	[ɕo]	[ɕɯ]
	[t]	[ta]	[te]		[to]	[tɯ]
	[ʧ]	[ʧa]		[ʧi]	[ʧo]	[ʧɯ]
	[ʦ]					[ʦɯ]
	[d]	[da]	[de]		[do]	
	[n]	[na]	[ne]		[no]	[nɯ]
	[ɲ]	[ɲa]		[ɲi]	[ɲo]	[ɲɯ]
	[h]	[ha]	[he]		[ho]	
	[ç]			[çi]		
	[ɸ]					[ɸɯ]
	[b]	[ba]	[be]		[bo]	[bɯ]
	[bʲ]	[bʲa]		[bʲi]	[bʲo]	[bʲɯ]
	[p]	[pa]	[pe]		[po]	[pɯ]
	[pʲ]	[pʲa]		[pʲi]	[pʲo]	[pʲɯ]
	[m]	[ma]	[me]		[mo]	[mɯ]
	[mʲ]	[mʲa]		[mʲi]	[mʲo]	[mʲɯ]
	[j]	[ja]			[jo]	[jɯ]
	[ɾ]	[ɾa]	[ɾe]		[ɾo]	[ɾɯ]
	[ɾʲ]	[ɾʲa]		[ɾʲi]	[ɾʲo]	[ɾʲɯ]
	[ɰ]	[ɰa]			[ɰo]	
	[N]	[N] No vowels				
	[Q] glottal stop	[Q] No vowels				

Japanese is a mora-timed rather than syllable-based language(Cutler and Otake, 1994) and this was the basic acoustic-motor representation used in the model. Mora are a sub-syllabic spoken unit in Japanese, which is typically formed by crossing one of the 29 consonants and one of the five vowels (see, *the table above*, summarized from Japanese language dictionary (Matsumura, 2006)). Several mora comprise just a consonant or a vowel – e.g., glottal stops [Q] and nasal consonant [N] do not have a vowel, and each of five vowels can form a mora without a consonant. Also, a long vowel consumes an additional mora: for example, *kyōto* is pronounced /kʲo:to/, in which case there are three moras. Although the exact number of moras is debated, the list above summarises the most widely accepted mora patterns in Japanese. The blank cells indicate where a specific combination of consonants and vowels do not exist as a real Japanese mora.

Item-selection

Table. Word-frequency and number of words in each cell of the training set

		Accent types		
		Flat-words	Type-1 words	Type-2 words
High frequent words	Log-10 Word frequency			
	Mean	3.58	3.58	3.58
	SD	0.44	0.46	0.54
	Number of words	500	330	25
	Frequency per million	Mean: 33.5, *SD*: 65.1, Range: 4.2-628.5		
Low frequent words	Log-10 Word frequency			
	Mean	1.79	1.80	1.80
	SD	0.72	0.72	0.72
	Number of words	500	330	25
	Frequency per million	Mean: 0.7, *SD*: 0.9, Range: 0.0008-4.09		

Japanese non-homophone, tri-mora nouns were used in this study. Shorter words were avoided because a very high proportion of them are homophones. Any homophone items from the tri-mora corpus were removed from the corpus as were seven words that contained a loan sound (e.g., [v]). Tri-mora words are also a useful length in that they clearly represent examples of time-varying word stimuli (see Main Text). From this word cohort, we selected 1710 training words while taking into account the following two variables: *word-frequency* and *pitch accent types*. The details are described below.

Word-frequency was obtained from the NTT Japanese psycholinguistic database(Amano and Kondo, 1999; Amano and Kondo, 2000; Sakuma et al., 2005), derived from the *Asahi* newspaper over a 14 year duration. This large sample contains approximately 240 million words[1]. Words that appeared just once (i.e., raw count word-frequency = 1) were excluded (404 words).

For tri-mora words in length, there are four types of pitch accent: *flat* words, *type-1* words, *type-2* words, and *type-3* words(Kindaichi, 2001). Type-3 words were excluded from this study as they were too rare to include. We selected high-frequent (HF) and low-frequent (LF) items from the remaining three accent categories so that the ratio of each accent type in the training set was identical to the entire NTT database (Ueno et al., In preparation). The table above summarizes the properties of the 1710 words included in the training set. The remaining 3511 tri-mora nouns in the database were used for testing generalization.

[1] We are extremely grateful to Tadahisa Kondo (NTT communication Ltd, Japan) for providing this information.

Representations

Twenty units in the input/output sound layers were dedicated to represent the distributed phonetic encoding of each mora and one unit was used to code the pitch accent types. Thus, a sequential presentation of three vectors (each, 21-bit) allowed us to represent every tri-mora word in the training corpus, uniquely. The specific coding are summarised below.

Table. Phonetic coding of each mora

Consonants

	Cons	App	Son	Cont	Strid	Nas	Voi	Asp	Glot	Ant	Dist	Lab	High	Low	Back
[k]	1	0	0	0	0	0	0	0	0	0	0	0	1	0	1
[kʲ]	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0
[g]	1	0	0	0	0	0	1	0	0	0	0	0	1	0	1
[gʲ]	1	0	0	0	0	0	1	0	0	0	1	0	1	0	0
[s]	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0
[ʃ]	1	0	0	1	1	0	0	1	0	0	1	0	1	0	0
[ʣ]	1	0	0	0	1	0	1	0	0	1	0	0	0	0	0
[ʤ]	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0
[t]	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
[ʧ]	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0
[ʦ]	1	0	0	0	1	0	0	0	0	1	0	0	0	0	0
[d]	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0
[n]	1	0	1	0	0	1	1	0	0	1	0	0	0	0	0
[ɲ]	1	0	1	0	0	1	1	0	0	0	1	0	1	0	0
[h]	0	1	0	1	0	0	0	1	0	0	0	0	0	1	0
[ç]	1	0	0	1	0	0	0	1	0	0	1	0	1	0	0
[ɸ]	1	0	0	1	0	0	0	1	0	0	0	1	0	0	0
[b]	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0
[bʲ]	1	0	0	0	0	0	1	0	0	0	1	1	1	0	0
[p]	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0
[pʲ]	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0
[m]	1	0	1	0	0	1	1	0	0	0	0	1	0	0	0
[mʲ]	1	0	1	0	0	1	1	0	0	0	1	1	1	0	0
[j]	0	1	1	1	0	0	1	0	0	0	1	1	1	0	0
[ɾ]	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0
[ɾʲ]	1	1	1	0	0	0	1	0	0	0	1	1	1	0	0
[ɯ]	0	1	1	1	0	0	1	0	0	0	0	1	1	0	1
[N]	1	0	1	0	0	1	1	0	0	0	0	0	0	0	1
[Q]	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0

Vowels	Voi	Rd	High	Low	Back
[a]	1	0	0	1	1
[e]	1	0	0	0	0
[i]	1	0	1	0	0
[o]	1	1	0	0	1
[ɯ]	1	0	1	0	1
No vowel	0	0	0	0	0

Abbreviations: Cons = consonantal; Apprx = approximant; Son = sonorant; Cont = continuant; Strid = strident; Nas = nasal; Voi = voiced; Asp = aspirated; Glot = glottalized; Rd = labial round; Ant = Coronal anterior; Dist = Coronal distributed; Lab = labial; High = Dorsal high; Low = Dorsal low; Back = Dorsal back

Phonetic coding of each mora.

Each mora was represented as the concatenation of the corresponding consonant and vowel(Matsumura, 2006), each of which can be coded in terms of distinctive phonetic features. The table above (*phonetic coding of each mora*) summarises the distinctive phonetic features of each consonant and each vowel in Japanese (Halle and Clements, 1983). Note that, other languages use additional features to define each phoneme (e.g., long vowels) but these are redundant in Japanese and so were not included in these simulations. As a result, each Japanese consonant was represented by a vector of 15 distinctive features and every Japanese vowel by a vector of 5 bits.

Pitch accent coding of each mora.

An additional unit was dedicated to code the pitch accent pattern of each word. Specifically, this unit was turned 'on' when a high pitch accented mora was processed (see the table below). Thus, in the case of flat words, this unit was turned 'on' when the 2nd and the 3rd moras were presented to the input layer or when these two moras were generated from the sound output layer. The same was true to the 1st mora of type-1 words and to the 2nd mora of type-2 words, respectively.

Table. The 'on'/'off' status of one of the units in the input/output sound layer as a function of mora position and accent types

	Accent types		
	Flat-words	Type-1 words	Type-2 words
1st mora	Off	On	Off
2nd mora	On	Off	On
3rd mora	On	Off	Off

Semantic representations

Semantic representations were abstract vector patterns of 50 bits. Each exemplar pattern was generated from 50 prototype patterns (Plaut and Kello, 1999; Rogers et al., 2004) in the following way. Each 'prototype' patterns were created by randomly setting 20/50 units to a value of 1 and 40 exemplars were generated from each prototype by randomly setting 10/20 'on' units to a value of 0. This procedure ensured that the 40 exemplars within the same prototype were relatively similar with each other (i.e., shorter Euclidian distance in the 50-bit multi-dimensional vector space) and relatively dissimilar with the exemplars from other prototypes (i.e., longer Euclidian distance). The resultant 2000 exemplars were checked to ensure that every pattern was different in the 'on'/'off' status of at least four units to all the other semantic patterns. These resultant semantic patterns were randomly assigned to one of the 1710 auditory-motor patterns to ensure the arbitrary mapping between the two representations of each word.

Lesioning procedure (cf. Figure 3)

Items

One hundred and twenty items were extracted from the training set with an orthogonal manipulation of word-frequency and accent types. The details are shown in the table below.

Table. Properties of the test corpus (20 items per cell)

		Accent types		
		Flat-words	Type-1 words	Type-2 words
Log-10 transformed word frequency				
High frequent words	Mean	3.633	3.542	3.681
	SD	0.525	0.365	0.559
	Per million	Mean: 33.4, *SD*: 66.8, Range: 4.4-379.3		
Low frequent words	Mean	2.287	2.142	2.126
	SD	0.510	0.591	0.371
	Per million	Mean: 1.1, *SD*: 1.2, Range: 0.06-4.09		
Spoken word imageability				
High frequent words	Mean	4.824	4.774	4.473
	SD	0.926	0.744	0.826
Low frequent words	Mean	4.850	4.888	4.659
	SD	0.946	1.061	1.038
Spoken word familiarity				
High frequent words	Mean	5.905	5.913	5.800
	SD	0.417	0.341	0.698
Low frequent words	Mean	5.566	5.459	5.326
	SD	0.411	0.899	0.827

Note: These values were taken from the NTT database(Amano and Kondo, 1999; Amano and Kondo, 2000; Sakuma et al., 2005).

Assessment

For all locations of damage, the model was lesioned 50 times at each of 15 levels of increasing severity and the 50 observations were averaged at each level of severity. This procedure ensures that a stable estimate for the effect of damage is extracted (Rogers et al., 2004).

Semantic inputs for speaking/naming

During training, the current (developing) semantic representation was used as the input to speech output (see *Method*). The status of semantic representations in brain-damaged patients is more complex. This is because having acquired a ‘matured’ semantic representation, even patients with a modality-specific impairment (e.g., word deafness) can

support their comprehension via input from other (intact) input modalities(Rogers et al., 2004). Therefore in speaking/naming, we presented an equal blend of (1) the current auditory-comprehension output under lesioning and (2) the matured comprehension output before damage (a proxy of the information activated by other non-speech inputs to semantics). Specifically, the latter was the output vector (50-bit) in vATL layer at the 3rd tick of comprehension, which was measured in the end of training. The former was the same output vector (50-bit) measured at every iteration of the lesioning-testing procedure (see above). From a different view point, this is a variant of *soft-clamping* (http://tedlab.mit.edu/~dr/Lens/inputTypes.html): The *external input* is the matured comprehension output, and the *initial output* was the comprehension output under lesioning. These two vectors were averaged with the *clamp-strength* of 0.5. Then, the resultant vector pattern was hard-clamped to the vATL-input layer whole the three ticks of speaking/naming.

Simulation of SD patients (Figure 3, Panel F)

There were two different procedures for simulation of SD patients: the nature of the degraded input to speaking and the application of damage. SD patients have multimodal, receptive and expressive semantic impairments in the context of anterior temporal atrophy (Lambon Ralph et al., 2001; Patterson et al., 2006; Patterson et al., 2007). Unlike patients with word deafness or agnosia, therefore, the SD patients have a breakdown of core semantic representations. Accordingly, for this patient group, we used the output of auditory comprehension as a proxy to the status of the semantic representations and used this same pattern as the input to speech production. In addition, Galton et al.(Galton et al., 2001), found that the atrophy in SD patients covers the ATL and temporopolar cortex in a particular distribution across the specific subregions. With regard to the two subregions included in our simulation, Galton et al found that the majority of atrophy occurred in the vATL whilst approximately only half of this rate of volume loss was found in the aSTG. Accordingly, we applied damage in the model to both regions but scaled them (2:1) in line with the neuroanatomical findings.

Simulation of Wernicke-type aphasia (Damage in Primary Auditory input layer. Figure 3, Panels E)

As noted in the main text, Wernicke-type aphasia was simulated by damaging the primary auditory input layer. Because this is an input layer over which the input is hard clamped, it does not have input connections like the other layers and damage tends to have less of an effect (with a hard-clamped input). We therefore adjusted the protocol for simulated damage by removing outgoing connections rather than ingoing connections.

Mixture of damage (Figure 3, Panels B & D)

Mixture of damage was achieved by simultaneously damaging two adjacent layers (Panel B: Insular-motor & iSMG layers; Panel D: iSMG & Primary auditory layers). Each of the

two layers were damaged in the same way as the pure damage case on each layer (e.g., Panels A, C, or E) with the level of damage applied equally across both sites.

Simulation of recovery (cf. Figure 4)

Items for assessment

Nonwords were the same 108 high-bimora-frequent nonwords used in Tanida et al. (see normative human experiments in Supplementary Data 2, S2). Real words were taken from the 60 low-frequent set used in the previous lesioning simulation (Figure 3). From these 60 words, nine items were removed in order to match their bi-mora frequency (Mean = 129668, *SD* = 166848) with that of the nonword items (Mean = 149009, *SD* = 28066).

Recovery period and learning parameters.

Human development in these simulations was captured by allowing the model to gradually update its weight structure, ultimately leading to normal 'adult' language performance. The same process was used to simulate spontaneous recovery post acute brain damage, by allowing the model to update the remaining weight structure in order to re-optimise performance (given its reduced computational resource post damage). This process uses the same learning algorithm in the developmental and recovery phases of the model. As such, the parameters (e.g., learning rates, weight decay, etc.) used in the simulation in the recovery phase were identical with those at the end of initial training (200^{th} epoch).

Analysis of the encoded similarity structure within each computational layer (cf. Figure 5)

The purpose of these analyses was to explore the nature of the similarity structure coded in the model. This was achieved by computing the similarity of the activation vectors at each target location (standardised Pearson's correlation coefficient), for all possible pairs of representations coded by the model (i.e., 1710*1709/2 observations). We then explored whether these 'observed' similarities were best predicted by semantic or acoustic-phonological similarity using multiple regression. The analyses were conducted with *Matlab* (http://www.mathworks.co.uk/).

Acoustic-phonological similarity (regressor)

This was computed as the standardized Pearson's correlation coefficient between the vectors representing the acoustic-phonological input patterns (21-bit × 3 mora positions). As noted in the Main Text, this was presented sequentially during input and so in order to compute a single correlation value, we treated the representation as a static single input by combining the three correlations into one (three Pearson's correlation coefficients were estimated between the 21-bit mora vectors and then these three coefficients were averaged). These values were standardized across all the pairs before entering into the regression model.

Semantic similarity (regressor)

This was the standardized Pearson's correlation coefficient between the vectors for the target semantic representations (50-bit). The estimated coefficients were standardized before entering into the regression model.

Internal representational similarity (criterion variable).

The dependent variable was the standardized Pearson's correlation coefficients between the unit activation vectors in each layer (e.g., 650-bit vector for aSTG layer). The activation pattern was measured at the 3rd tick of comprehension (this is identical with the pattern in the 3rd tick of repetition) because at this point the entire word had been presented across the input layer. The only exception to this was the similarity structure computed for the activations observed at the primary-auditory input layer. We used a slightly different procedure for this site because the activation pattern of this layer at the 3rd tick reflected just the hard-clamped 3rd mora input and not the entire word. Therefore, to obtain an estimate of the similarities encoded at this input layer, we 'concatenated' the three 21-bit output vectors (one for each time point: 1st, 2nd, & 3rd time tick). We should note here that it is not surprising that the acoustic-phonological-similarity regressor strongly predicted this dependent variable, but it was included in order to show the maximum possible acoustic-phonological similarity encoded within the trained model and to use this as a comparator for other, more variable representational layers.

Speaking/naming (Figure 5b in the main text)

Figure 5b captures the shift in sensitivity to semantic and acoustic-phonological similarity across successive representation layers during speech output. We measured the unit activations in each layer at the 3rd tick/mora output of speaking. At this point, the network had already produced the 1st and the 2nd mora at the output layer and these elements of the output had already been fed-back to the previous layers.

Semantic speaking/naming errors (cf. Figure 6)

Item for assessment

Eighty-three, close semantically-related word pairs (166 items) were extracted to test the rate of semantic speaking/naming errors after different locations of damage in the model. These items were the closest possible semantic neighbours in the sense that their distributed semantic representations differed by only four 'on' bits (the minimum allowed when the representations were computed – see above). High frequency items were used as the probe as they activate the units in the semantic system most strongly.

Lesioning

Four internal layers were separately damaged using the same procedure described above (Figure 3). To increase the power of the analysis, lesioning was conducted twice for each of the five simulations with different starting random seeds (generating 10 observations). The severity of damage (the combination of noise and link removal) was scaled for each layer such that the levels of speaking/naming accuracy were the same irrespective of lesion location. This then allowed us to compare the rate of semantic speaking/naming errors, independently of the level of speaking/naming impairment. A similar process is commonly used in aphasiological studies and the target voxel-symptom lesion mapping study (to correct for overall levels of accuracy) (Schwartz et al., 2009) . The table below summarises the specific levels of damage at each location in order to equate overall levels of speaking/naming accuracy. A 4 (lesion location) × 4 (severity level) ANOVA on speaking/naming accuracy (Figure 6a in the main text) confirmed the accuracy matching: the simple effect of lesion was not significant at any level of severity (all $p > .15$).

Table. Levels of damage used to achieve matched naming accuracy across different locations of damage.

Severity	Lesion parameters	Damaged layers			
		iSMG	mSTG/STS	aSTG/STS	Tri-Ope.
Mild	Range of noise	0.096	0.0276	0.026	0.047
	Proportion of link removal	0.048	0.0138	0.013	0.0235
Moderate	Range of noise	0.132	0.036	0.032	0.058
	Proportion of link removal	0.066	0.018	0.016	0.029
Moderate-severe	Range of noise	0.192	0.0484	0.041	0.076
	Proportion of link removal	0.096	0.0242	0.0205	0.038
Severe	Range of noise	0.312	0.068	0.056	0.104
	Proportion of link removal	0.156	0.034	0.028	0.052

Semantic errors (Figure 6b in the main text)

Semantic errors were evaluated based on the *best-fit criterion* in the multi-dimensional vector space. Specifically, the output vector in speaking (21-bit * 3 time ticks) was compared to every auditory pattern in the 1710 training set and Euclidean distance was calculated (the distance was calculated per time tick/mora and three distances were averaged). If the observed output and the semantic neighbour produced the smallest Euclidean distance (i.e., best fit), then the output was scored as a semantic error. In each condition, the lesioning and testing procedure was reiterated 50 times and the performances were averaged.

Supplementary References

Amano, S., and Kondo, T. (2000). Japanese NTT database series: Lexical properties of Japanese, word-frequency (11) (Tokyo, Sanseido).

Cutler, A., and Otake, T. (1994). Mora of phoneme - Further evidence for langauge-specific listening. Journal of Memory and Language *33*, 824-844.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1996). Rethinking Innateness: A connectionist perspective on development (London, MIT Press).

Hinton, G. E. (1989). Connectionist learning procedures. Artificial Intelligence *40*, 185-234.

Kindaichi, H. (2001). In Shinmeikai nihongo accent jiten (Japanese accent dictionary), K. Akinaga, ed. (Tokyo, Sanseido).

Matsumura, A., ed. (2006). Daijirin (Japanese disctionary), 3rd edn (Tokyo, Sanseido).

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. Language and Cognitive Processes *12*, 765-805.

Plaut, D. C. (2002). Graded modality-specific specialisation in semantics: A computational account of optic aphasia. Cognitive Neuropsychology *19*, 603-639.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., and Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. Psychological Review *103*, 56-115.

Sakuma, N., Ijuin, M., Fushimi, T., Tatsumi, I., Tanaka, M., Amano, S., and Kondo, T. (2005). Japanese NTT database series: Lexical properties of Japanese, word-imageability (Tokyo, Sanseido).

Tamaoka, K., and Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. Behavior Research Methods Instruments & Computers *36*, 531-547.

Ueno, T., Saito, A., Saito, S., Patterson, K., and Lambon Ralph, M. A. (In preparation). Not lost in translation: The impact of semantics in Japanese phonological short-term memory.