The Co-Ascription of Ordered Lexical Pairs: a Cognitive-Science-Based Semantic Theory of Meaning and Reference: Part 2

Tom Johnston
tm1972.com
academia.edu
tmj44x@att.net
August, 2019

Table of Contents

Abstract

This is Part 2 of the semantic theory I call TM. In Part 1, I developed TM as a theory in the analytic philosophy of language, in lexical semantics, and in the sociology of relating occasions of statement production and comprehension to formal and informal lexicographic conclusions about statements and lexical items – roughly, as showing how synchronic semantics is a sociological derivative of diachronic, person-relative acts of linguistic behavior. I included descriptions of new cognitive psychology experimental paradigms which would allow us to precisely measure the two constituents of semantics – meaning and reference – both at the level of individual speech acts and at the level of societal convergences, i.e. at both the token and type levels.

In the Introduction, I recapitulate the arguments of Part 1. The Introduction also develops some analytic philosophical and lexical semantics themes not discussed in Part 1.

After the Introduction, I present neural TM (nTM) as a theory of the neural mechanisms and processes which give rise to these person/occasion-relative acts of linguistic behavior. I develop nTM at three levels, the first two of which describe linguistic/semantic functions independently of their cortical locations. At the first level, I describe individual word-to-word and word-to-object connections. At the second level, I describe the corresponding structuralist networks of which they are the individual components. At this level, I introduce some key linguistic concepts of TM – its graded meaning, reference, and generalization sets, and the types of statements which express various levels of word-to-word and word-to-object relationships among lexical items which, because of the constraints they impose on the use of those lexical items in statements we produce and comprehend, are concepts. This constitutes the second structural level of nTM.

At the third structural level, I associate the non-localized structures of the previous levels with cortically located neural structures and the fasciculi that connect them. I distinguish neural areas in which primary (phonetic) and secondary (orthographic) lexicons are stored in long-term memory. I also describe the *embodied concepts* which co-exist in the anterior temporal lobes with the images they lexicalize. These concepts are often said to name physical objects and their features, although what they in fact name are *kinds* of physical objects and features. I describe how conceptual constraints and referential constraints interact to channel our intentions to say how things are into statements which are semantically well-formed, and which consequently successfully communicate information.

Following this presentation of nTM, I examine five prominent neural semantic theories. I point out what is wrong with each of them as far as their explanations of semantics are concerned, and I also indicate how nTM can replace the "semantic cores" of those theories.

The two basic mistakes made by neuroscience semantic theories, as I will explain, are (i) that all but one of them regard semantics as a matter of the association of words with perceptual images, and generalizations from those associations; and (ii) that they all rely on an unspecified set of neural structures which purportedly encode the meaning of concepts in abstraction from their phonological and orthographic forms. nTM maintains, in contrast, that there are no abstract neural representations of semantic content. Neural constraints on our linguistic behavior, especially on our ascriptive and co-ascriptive use of words, express the semantic constraints on those words which make them concepts. That *is* the semantic content of words.

I next consider several results from neuroscience experimental data which have been given one interpretation by one or another of the standard neurosemantic theories, but to which nTM gives a different interpretation. I include several predictions which I have found neither confirmed nor disconfirmed in the experimental neuroscience literature.

After a concluding section in which I summarize the major changes to neurosemantic theory introduced by TM, and the analytic philosophy of language and lexical semantics contexts within which TM is situated, there follows an appendix in which I discuss neural net AI, and make some recommendations for implementing nTM in silicon.

Introduction

Let me summarize TM as it has so-far been presented. This summarizes the perspectives from all the boxes in Figure 1 below except the final one -- the neuroscience box. This Figure was introduced in Part 1, where a more complete account was given of the framework it provides for TM.

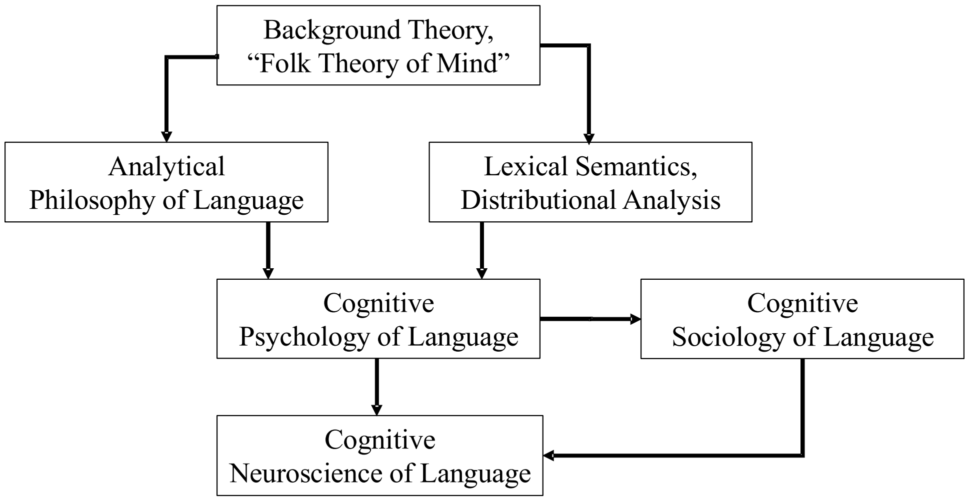

Figure 1. Theoretical Perspectives on Language.

Semantics consists of the mechanisms and processes by which language conveys information. According to TM, we convey information by making statements, and we make statements by picking some thing out and then saying something about it. The *basic statements* which are the focus of TM are simple declarative sentences, produced and comprehended in a context within which they represent, and are assumed to represent, what the speaker believes to be true. I focus on basic statements in order to exclude syntax from this study of semantics, and because I believe that the information content of any statement, no matter how complex, can be expressed primarily as a boolean combination of basic statements.

As manifested in reference, semantics is the association of a sub-sentential linguistic unit of meaning, called a *lexical item*, with a perceived pattern. Our folk ontology tells us that these perceived patterns represent objects, or features (properties or relationships) of objects, or processes in which one or more objects participate. The neural association of a lexical item and what it refers to relates two perceptual patterns with one another, since a lexical item is itself neurally recognized and remembered in just the same way that any other perceivable thing is.

As manifested in meaning, semantics is the association of two lexical items in which the ascription of one of the lexical items to a referent influences the ascription of the other lexical item to that same thing, on that same occasion. If we call someone a bachelor, we cannot object to someone else saying that he an adult, and cannot agree with yet someone else who says that he is married. This co-ascription relationship between two lexical items manifests itself in this sort of linguistic behavior, and rests ultimately on Hebbian-learned evolving associations between the neural representations of each of the pair of lexical items occurring in the statements which include them.

The association of lexical items and basic statements is this. A basic statement consists of two lexical items.[1] One is used, together with an indexical, to pick something out as an instance of a kind of thing. The other is used to ascribe a property or relationship to that thing. Neuroscientists generally call this *naming* the thing or the feature it has, but it is more accurate to call it designating the kind, or category, of that thing or feature.

Basic statements are fully-instantiated statements. They contain no variables. As for statements which include variables, TM recognizes both universally-quantified (All X are Y) and existentially-quantified (Some X are Y) statements as records of patterns of co-ascriptional behavior generalized from basic statements produced on specific occasions.

Both philosophers and lexical semanticists have been attempting to distinguish statements true because of how the world is from statements true because of how language is. I take their lack of success as evidence for Quine's contention that there is no sharp distinction to be discovered, and instead that the distinction between the two kinds of statements – analytic and synthetic ones and, perhaps orthogonally, a priori and a posteriori ones – is a matter of degree. To bring word-pair co-occurrence data more directly to bear on the issue of lexical meaning, I recommend that lexical semanticists (i) focus on co-ascribed word-pairs rather than the broader category of textually near-neighbor word-pairs, and that they (ii) supplement statistical analyses of co-ascription patterns in linguistic corpora with a form of research that of necessity requires a language-using subject to make reports on his own state of mind, in particular on how strong his concept-to-object associations are, and how strong his concept-to-concept associations are. An experimental framework for this investigation was described in Part 1. From this description, it is clear how lexical semanticists and cognitive psychologists can set up concept ascription events to measure reference, and concept co-ascription events to measure meaning, querying their participants to obtain reports on the strength of these extensional and intensional relationships.

Once we reach the Figure 1 level of cognitive psychology, it is pretty much a foregone conclusion that there will be, indeed must be, a neuroscience explanation of the states and processes described at the higher levels shown in that figure. Roughly, cognitive psychology is the interface between the neuroscience level of Figure 1 and the other four

[1] As explained in Part 1, a lexical item may consist of more than a single word. However, I will frequently use the word "word" in this text, instead of the phrase "lexical item". In this essay, and in Part 1, "word" and "lexical item" are generally used synonymously.

levels shown there.

TM relies on a single mechanism at the neurophysiological level of explanation. At that level, Hebbian learning is that mechanism. The learned association of the physical wordforms for two lexical items is the basis for co-ascription constraints between the concepts they represent. The learned association of a wordform for a lexical item with any other perceived pattern is the basis for referential constraints between that concept and that noticed component of the external world. Both patterns are encoded, associated, remembered, and accessed in human brains.

Basic TM Terminology

In Part 1, I defined a *concept* as a sub-sentential unit of meaning expressed by a physical phonetic or orthographic string, the concept's *wordform*. I also defined a *basic statement* as a declarative sentence, one which expresses something that the speaker believes to be true, which contains two concepts. The first concept is the subject of the sentence, indicating what *kind* of thing some thing is.[2] The second concept is the predicate of the sentence, indicating a *feature* which that thing has. "That dog is black" says, of the object which is the *referent* of the statement, that it is an instance of the kind [Dog]. It is the kind and the indexical expression "that" which together pick out the referent. The statement also says, of its referent, that it possesses the feature [Black].

"All dogs are animals" is another kind of statement, one in which the concept [Dog] is both a kind and also the referent of the statement. These statements are what logicians call universally-quantified statements, and the word-pair associations they express are integral elements in the composite of word-pairs that constitute the meaning of the subject concept ([Dog] in this case).

What is to be Explained

In summary, TM includes the following claims, for which later sections will provide a neurophysiological explanation:

1. The semantics of language consists of the evolving mutual influences of meaning and reference, as pairs of concepts occur in statements. For each person and for each linguistic community, meaning and reference are realized, respectively, as Graded Meaning Sets and Graded Reference Sets (see Part 1, and below). Each such set, for each of us and at each point in time, exists as a group of neurons and their patterns of firing.

[2] In this essay, I use "some thing" and "something" as different lexical items. In the former case, "some" is a generalization word, and "thing" denotes a discrete object. In the latter case, "something" is used in its broader, ordinary language sense in which it can pick out a discrete object ("Something's hiding in the bushes over there"), a feature of an object ("There's something about Mary"), a relationship among objects ("Something keeps them together"), or even a process ("Something's going on in the next apartment").

2. We engage in acts of linguistic production and comprehension which manifest the constraints these two graded sets exert on one another, for each of us. The mutuality of meaning/reference constraints was explained in Part 1, and was there called the Meaning/Reference Reciprocity Thesis (MRRT).

3. By talking to one another, we provide examples for each other of ascriptive and co-ascriptive uses of concepts in statements. This talk is sensory input for each of us in which patterns of co-occurrence of the physical wordforms of concepts occur. Repeated perceived patterns of co-ascription are retained in long-term memory (LTM) as neural states that are modified on each occasion of use of those pairs of concepts. In our conversations (and written exchanges), our mutual exemplifications of meaningful language converge across a linguistic community – of you and I in a conversation, and of all of us in wider overlapping confluences. The driving force is the need to successfully exchange information, a force operative across evolutionary time, conversation-specific timeframes, and all timeframes in-between.

4. *Graded Meaning Sets* are sets of ordered pairs of concepts, the first concept helping to pick some thing out by saying what kind of thing it is, and the second concept saying something about that thing by pointing out a feature it instantiates. Meaning sets relate words to words. This is an ordering of co-ascriptional use of the two concepts in the same statements, and is not necessarily a temporal order in which those words occur in those statements. But picking something out semantically comes first; saying something about it semantically presupposes it.

5. *Graded Generalization Sets* are sets of unordered pairs of concepts, each one believed to have one or more referents that also instantiate the other concept. These can be represented as Venn Diagrams whose intersection of the two concepts is not empty. They encode much of our knowledge about the world.[3]

6. *Graded Reference Sets* are sets of word-to-object pairings, each one consisting of a physical wordform and another perceived pattern, that one being of the thing or feature the word refers to. Reference sets relate words to objects. At the neural level, sensory input is perceived as manifesting a pattern because the sensory input activates a long-term-memory (LTM) of similar images. Mid-way in sensory processing, in particular in temporal

[3] In Part 1, Generalization Sets were called Association Sets.

and intraparietal areas of the brain, single-modality sensory input becomes integrated into remembered images of external objects and features.

7. A *basic statement* refers to some thing, and says something about it. In the case of basic statements about concrete things and features, I call what is referred to a *lexicalized image* – or, more ontologically speaking, an object whose awareness activates a lexicalized image in the brain. As lexicalized, it is a conceptualized image. What is said about it is that it instantiates what a second lexicalized image denotes, that being an image of some kind of property or relationship that thing has.

Aside: here is an important point that I will return to later on. Monkeys, who lack language, can nonetheless recognize tigers, and can produce "words" for tigers, and different words for other predators. The image of the referent and the image of a word which designates what kind of referent it is, are neurally associated and recorded in the monkey's LTM. As impressive as the association of an image with a category of referent is, something more happens when language-using human beings associate words and images of things. For we human beings, those images become lexicalized images. Instead of just words for categories of things like tigers, human beings produce words for categories of things but these words are also constrained in their use by their associations with other words in lexicons.[4] These words are, because of these associations, full-fledged concepts. I will have more to say on this important point later.

Let us assume that the commonsense, analytic philosophical, and lexical semantics statements we make about language have been shown to be abstractions from the person- and time-relative occasions on which we produce or comprehend a statement. Let us assume further that these token-level linguistic phenomena have been shown to be the result of the mutual influence of Graded Meaning Sets, Graded Reference Sets, and Graded Generalization Sets, for each of us, and then for all of us collectively. These assumptions are among the conclusions reached in Part 1.

This alone provides a resolution of truth-functional and structuralist approaches to semantics, by establishing a cognitive psychology basis for both reference and meaning which addresses many of the issues which appear in philosophical and lexical semantics work on semantics. The last step, as indicated in Figure 1, is to provide a cognitive neuroscience account of this theory.

4 Convolutional neural networks are now pretty good at recognizing objects as instances of categories – at creating a neural representation of a scanned object, and then of correctly labeling it. They have pretty much caught up to monkeys. But, I propose, these neural networks are not able to use those labels as concepts, because current linguistic and neuroscience theory does not understand how remembered relationships among words constrain their use equipotently and interactively with remembered relationships between words and images.

Neural TM: Fundamental Hypotheses

TM, as an analytical philosophical and lexical semantics theory, was developed in Part 1. There I also described a cognitive psychology paradigm for isolating, within individual speech acts, the components of a word's meaning and of its reference, and the dynamics of the reciprocal influence which meaning and reference exert on one another.

I now develop a theory of the neurological foundations for these two components of semantics.

Four principles govern nTM – neural TM.

- <u>Physical Wordforms</u>. There are no mental or abstract representations of concepts in the brain. There are only neural representations of the physical wordforms we encounter in the form of phonetic images, orthographic images, or sign-language images.

- <u>Essentially Related Concepts</u>. Current neuroscience regards concepts as atomic. The assumption is that the meaning of each concept stands alone, independent of other concepts. TM rejects this. For TM, and for Saussurean lexical semanticists, words are connected to other words via a structuralist network of word-to-word associations. For TM, this network constrains the co-ascription of pairs of these words to the same referent, on the same occasion, and it is these constraints which, as Saussure first proposed, constitute their meaning. This network is manifested in our verbal behavior as patterns in the co-ascription of words to the same referents.

- <u>Lexicalized Images</u>. Wordforms for *embodied concepts* – corresponding to what neuroscientists call concrete words – are connected to images of the things they refer to via a network of word-to-image associations. This network is manifested in our verbal behavior as constraints on the ascription of words to the physical things and features we see around us. Embodied concepts are what I also call *lexicalized images* – each term emphasizing one of the two components, concept or image.

- <u>Semantics</u>. Reference, in the form of lexicalized images, and meaning, in the form both of embodied concepts and of abstract concepts, mutually constrain their ascriptions and co-ascriptions, resulting in the production and comprehension of semantically well-formed statements, ones that successfully convey information. This is TM's MRRT – the Meaning/Reference Reciprocity Thesis.

Meaning and Reference in Neuroscience

In a nutshell, neuroscience studies of semantics manifest a conceptual hemianopsia. With the exception of the Dual Coding Models of Paivio and Vigliocco, neuroscientists equate semantics with reference, i.e. with the relation of words to images, and with consolidations and generalizations of those relationships.[5] This focus can be traced to the meta-hypothesis of the Grounded Cognition Model, and from nearly universal agreement that the Amodal Symbolic Model (most prominently developed as Jerry Fodor's Language of Thought hypothesis) is inconsistent with neurological findings about the brain and language.

But since Immanuel Kant first presented the distinction between analytic and synthetic statements, and the related distinction between a priori and a posteriori statements, philosophers dealing with language have recognized that some statements are true or false regardless of what things happen to be like in the world we perceive around us. These are the a priori true statements, ones like “All cardiologists are doctors”, or the time-worn example “All bachelors are unmarried”.

Most, and perhaps all, such statements are thought to also be analytic statements, which are ones true solely because of the meaning of their constituent lexical items. “All bachelors are unmarried” is true, not because the world happens to contain no married bachelors, but, rather, is true because being unmarried is part of the meaning of being a bachelor. This is manifested as a constraint on predicative behavior, that once having said that someone is a bachelor, if we then went on to say that, nonetheless, that person is married, we wouldn't be thought to be deficient in our acquaintanceship with bachelors. We would be thought to show that we didn't know what “bachelor”, at least in its standard sense, means.

Definitions of lexical items are paradigmatically semantic phenomena. Of course, the reference of words that name kinds of perceivable things, or generalizations of the kinds of things they are, is also a semantic phenomenon. The two of them together, reference and meaning, the extension of our words as well as their intension, are equipotent in patterning our linguistic behavior so we can communicate information, one to the other.[6]

As the word “reciprocity” indicates, semantics isn't just the two – reference and meaning

[5] The Dual Coding Models are discussed later in this essay.

[6] The terms “intension” and “extension” originated with Carnap, and can be considered equivalent to Frege's terms “sense” (sinn) and “reference” (bedeutung). These terms also have a use in set theory, one very relevant to this discussion. A set is a collection of all and only those individuals in some universe of discourse who satisfy the membership criterion for the set. The collection is the extension of the set, the things that belong to it; the membership criterion is the intension of the set, the criterion for belonging to it. My use of the terms “meaning” and “reference” can be considered closely equivalent to the standard use of these two pairs of terms. Neuroscience use of the term “meaning” is loosely equivalent, unfortunately, to the standard philosophical and linguistic use of the term “reference”.

– sitting side by side. Semantics, manifested in intelligible statements made by each of us, is a combination of lexical items in statements which reflects an accommodation of each lexical item to the other, and each to what it refers to, in the mutually-influencing modalities of meaning and reference. If a statement is judged by others to be mistaken, it may be primarily because one or the other of the two lexical items in the statement are judged to not correctly name the kind of their purported referents. Or it may be primarily because the co-ascription of the two lexical items is judged to evidence a faulty understanding of what one or both of them mean. Usually, it is because there seems to be no satisfactory way of interpreting the statement, either by a situationally-specific acceptance of an extended referential range for either or both concepts, or by a situationally-specific acceptance of an extended sense of what either or both of them mean.

Since neuroscientists are hot on the trail of the reference component of semantics, they should have no trouble understanding TM's Graded Reference Set account of reference, initially presented in Part 1. But since they seem unable to see the semantic significance of word-to-word co-ascriptional constraints, whose strongest components are expressed in dictionary definitions of words, they may have difficulty in understanding the structuralist implications of TM's Graded Meaning Set account of meaning, that lexical items form a network of constraints on one another which constitutes the meaning of each of those items. This thesis was first put forward by Ferdinand de Saussure, and continues to be developed in the field of structuralist lexical semantics, and empirically investigated in lexical-pair co-occurrence studies.

Lexical semanticists have, however, failed to distinguish meaning and reference as separate sources of word-pair co-occurrence regularities in their studies of linguistic corpora. For example, “horse” and “pasture” often occur as word-pairs, and this is clearly because, in this world of ours, horses are frequently found in pastures. On the other hand, “horse” and “animal” also occur as word-pairs, but not because our perceptual experience shows us the things called horses exhibiting the quality of being animals, but rather because “animal” is part of the meaning of “horse”, just as being unmarried is part of what being a bachelor means.

Lexical semanticists have not been able to tease out these two different influences on word-pair co-occurrence. TM can, and it did so in Part 1, especially in the sections which developed a cognitive psychology experimental paradigm for both meaning and reference, and in the cognitive sociology accounts of how individual speech acts of making statements alter co-ascriptive patterns for pairs of lexical items that, if widespread enough over speakers and extents of time, eventually appear as adjustments to the definition of one of the lexical items in those pairs.

One synoptic expression of this difficulty is expressed as the *word-world distinction*. In this form, the question is how to distinguish statements which are true because of what their constituent words mean, regardless of what the world they may refer to is like, from statements whose truth is, given what those words mean, true because they correctly describe something about this world of ours.

Another synoptic expression of this difficulty is expressed as the *encyclopedia-dictionary distinction*. In this form, the question is how to distinguish statements which belong in an encyclopedia from statements which belong in a dictionary. Encyclopedias contain statements that describe what things in the world are like, statements that can be falsified if things are discovered to not be like that, on closer inspection. But what things are like in the world does not falsify the definitional statements found in dictionaries. Those statements are true no matter what; denying their truth evidences ignorance of what words mean, not ignorance of what the world around us happens to be like.

But all of this was explained, in much more detail, in Part 1. There, Kant's distinctions were traced through their development from the Logical Positivists of nearly a century ago and through the canonical statement of logical positivism in Wittgenstein's *Tractatus* (which he immediately began to deconstruct) and in Carnap's *Aufbau* (which Quine began to deconstruct in his seminal “Two Dogmas of Empiricism” and in subsequent pursuit of his program of naturalizing epistemology). The analytic/synthetic and a priori / a posteriori distinctions still exercise contemporary analytic philosophers, perhaps because a central thesis of Quine's, that the analytic/synthetic distinction is a matter of degree, with statements falling along a continuum between those limit concepts, has never been adequately cashed in. TM, and especially its Graded Meaning Set component and the manifestation of that component in speech acts, does cash it in.

So the point is this.

Meaning is a structural network of constraints which lexical items place on one another. Lexical semanticists, following in Saussure's structuralist tradition, try to articulate this network. According to Lenci (2008, 2014), they have so far failed. TM, I maintain, succeeds where they have not.

Reference is a naming relationship between words and the kinds of things they refer to. It is what accounts for the fact of *aboutness*, that what we say is ultimately accountable to what the world we are aware of is like. Analytic philosophers of language, following in the truth-functional semantics tradition of Carnap, Montague, Davidson, and Lewis, try to articulate this relationship.

Barbara Partee famously proposed that we might never be able to find a single subject matter that unites these two approaches.[7] As I claimed at the conclusion of Part 1, TM does show how truth-functional semantics and structuralist semantics are indeed about

7 See the discussion in (Santambroglio & Violi) 1988, and the reference to Partee on p.13.

“the same thing”.

I also note that nTM has a narrower focus than the neural semantic theories I will compare it with, and this focus can be expressed like this: nTM is a theory of a phonological and an orthographic neural lexicon, of the connection of words to the perceived images of things, and of the processes which form statements from this joint repository of lexical and object recognition material. But it is not a theory which traces the production of statements to motor output, or the comprehension of statements from original sensory input, or of the ability to replicate verbal utterances.

The First Structural Level of nTM

nTM has a structure and set of associated functions which I will initially describe in abstraction from their localization in areas of the human brain. I begin by describing a concept, which is the unit of meaning for the first wordform in an ordered pair of wordforms. Ordered word-pairs are the units of *Graded Meaning Sets*, and each set, for the first lexical item in the pair, fully accounts for the meaning of that lexical item – which, because of those connections, manifests itself as a full-fledged concept. Unordered word-pairs are the units of *Graded Generalization Sets*, and each such unordered pair represents a statement which, in a normal Gricean context, expresses a generalization which the speaker believes to be true about the world, and which the audience accepts as such an expression. And finally, embodied concepts and the referents they name are associations constituting *Graded Reference Sets*, and each such pair associates a concept with the image it lexicalizes.

I begin with word-pairs, the components of the first two graded sets.

Word-to-Word Connections

In its minimal physical and semantic form, a concept is an ordered pair of words. In its full form, it is a set of such ordered pairs, including the network extending out from the second words in those pairs, ultimately relating all concepts to all other concepts. In the strongest form of connection between those concepts, their co-ascription to the same referent cannot be denied without thereby committing a semantic mistake. I will call this an *entailment connection*, because the ascription of the first word to a referent entails the ascription of the second one. For example, [Bachelor::Unmarried] is such an ordered pair, so given that a referent is a bachelor, to deny that he is unmarried is to reveal that one does not understand the ordinary and widespread meaning of the word "bachelor". Being a bachelor entails being unmarried.

> Aside: I use "word-pair" in a narrower sense than the one used by lexical semanticists. For them, a word-pair is any pair of lexical items occurring in close proximity in a stretch of text. For me, word-pairs are pairs of lexical items that are co-ascribed to the same referent on the same occasion. For purposes of study, we may concentrate on word-pairs which occur within the same statement. But any pair of lexical items ascribed to the same object (on the same occasion), is a word-pair. Each lexical item puts the object into the category it names, and so a word-pair could be represented as a Venn Diagram, with the referent objects located in the intersection of the two depicted sets.

The minimum entailment connection is a word-pair of the form shown in Figure 2.

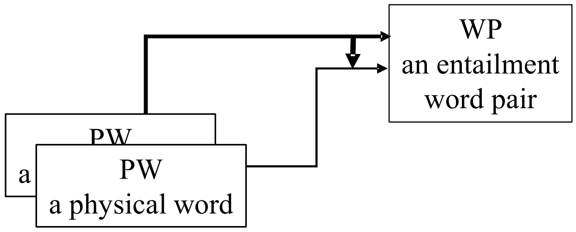

Figure 2. An Entailment-Strength Word-Pair.

This and all but one of the following five figures show two representations, each of a different word, and their association as a word-pair. In this Figure, the thick arrow represents the first word in an ordered pair of words, and the thin arrow the second word. The short arrow from the first PW/WP link to the second one is also a thick arrow. It indicates that the co-ascriptional relationship is an entailment-strength relationship, one whose violation evidences a misunderstanding of the meaning one or both concepts represented by the two physical words.

I interpret a misunderstanding of meaning, in linguistic behavioral terms, as manifested in being able to imagine a counterexample to a co-ascription which the rest of the linguistic community cannot imagine.[8] The linguistic community, consequently, will regard the purported counterexample as a misdescription, as describing nothing – real or possible – at all. There is no such thing as a horse which is not an animal; and no matter how much the world might change, there never could be such a thing. That is, the rest of us can't imagine such a thing.

If we ever could correctly describe some thing as a horse which was not an animal, the meaning of either or both of those two words would have changed, in a very fundamental way. Either or both phonetic/orthographic strings would then represent different concepts.

How are these word-pair associations formed? What creates the relationship by which, given the ascription of the first word to something, withholding the ascription of the second word is a semantic mistake, and by which any purported counterexample is in fact an example of neither anything real nor of anything conceivable? Since at this first level of nTM, we have abstracted from its neural implementation, the only thing we can say is that it is a learned association. We learn that we can't apply the first word to anything and at the same time deny that the second word also applies, because if we tried to do that, the consensus pattern of co-ascription within our linguistic communities would lead other fluent speakers to assert that whatever example we proposed would describe neither anything in the real world or, indeed, anything in any scenario we could coherently describe.

Another category of word-pairs is shown in Figure 3.

[8] See Part 1, pp.23-25. Also (Lyons, 1995), pp. 142, 188, 282.

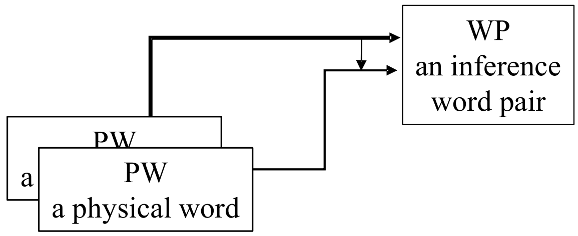

Figure 3. An Inference-Strength Word-Pair.

In this Figure, as in Figure 2, the thick arrow represents the first word in an ordered pair of words, and the thin arrow the second word. But in this Figure, the short arrow from the first PW/WP link to the second one is a thin arrow. It indicates that the co-ascriptional relationship is an inference-strength relationship, one whose violation (in a purported counterexample) evidences, not a misunderstanding of the meaning of one or both concepts represented by the two physical words, but rather the mistaken belief that there exists something which instantiates the first concept, but not the second one.

> Aside: the terms "entailment" and "inference" are used, by logicians, to distinguish "if...then" statements which are necessarily true from those which are contingently true. In modal logic, this expresses the difference between if...thens which are true in all possible worlds from if...thens which are true only in some possible worlds, such as this actual world. I use these terms in the same way.

These word-pairs express what we believe to be universally true about the actual world. For example, someone might describe something as a salamander that can exhale very hot gas, but most of us would be quite certain that there is no such creature. It's not that the description is semantically incoherent. For most of us, it is not. Most of us can imagine such a creature (or, in a Tolkien world, its larger cousin, a dragon). It's just that we all know that there isn't anything like that in the real world. [Salamander::Not able to exhale hot gas] is universally true in this world we live in[9,10].

How are these word-pair associations formed? What creates the relationship in which, given the ascription of the first word to something, withholding the ascription of the

9 For others of us, particularly scientists who study such things, the connection between "salamander" and "not able to exhale hot gas" might be an entailment-strength connection, something to which they can imagine nothing, no matter how outre, to which they would apply the first word but withhold the second one.

10 In Part 1, this distinction between entailments and inferences was not emphasized. Instead, I made the Quinean point that true universally-quantified statements fall along a semantic continuum from analytic to synthetic, and along an epistemological continuum from a priori to a posteriori, and that for both continua, the end points are like the limits in calculus, to be approached but never instantiated. It follows that entailments and inferences are not pure, sometimes instantiated, forms, but rather a matter of what kinds of judgments we will make, on various occasions, on why a co-ascription is correct or incorrect. Here in Part 2, I have found it helpful to make greater use of the distinction between entailments and inferences, even given that the distinction is a matter of degree in patterns of verbal behavior, and not a dichotomy.

second word reveals a factual misunderstanding, incorrect knowledge about some aspect of the real world? Again, these word-pairs are formed because they too are learned associations. We learn that if we apply the first word to something and at the same time deny that the second word also applies, we will be judged by knowledgeable speakers to be factually mistaken. Any such purported example of something that really exists will be perfectly comprehensible. No semantic mistake will be imputed, simply the judgment that we have a piece of incorrect knowledge about the world. There happen to be no such things.

A third category of word-pairs is shown in Figure 4.

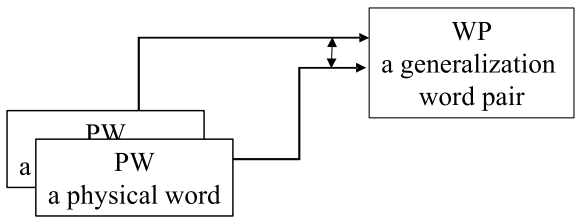

Figure 4. A Generalization-Strength Word-Pair.

In this Figure, unlike in both Figures 2 and 3, the first and second word-to-word-pair links are of equal thickness, indicating that the word-pair is not ordered. This represents a co-ascriptional relationship between two words that is a generalization rather than a universally valid co-ascription. A counterexample need evidence no misunderstanding of the meaning of either or both concepts represented by the two physical words, nor a violation of the referential use of those words. The short arrow between the two PW/WP links is bi-directional, indicating that, given an ascription of either concept, co-instantiations of the second concept will be acknowledged to exist, but not in all cases.

These generalizations claim that there are some (or many, or few) co-instantiations of the two word-pairs, but not that instantiations of one but not the other do not exist. On the contrary, these generalizations simply claim that there are such things, that such things really do exist. These are the things to which both words apply. “Some dogs are hairless” is a true generalization from the rather uncommon occurrence of hairless dogs. “Some dogs have fur” is a true generalization from the rather common occurrence of dogs with fur.

These word-pair associations are also learned. And they are semantically linked to the first two categories, the ones governing universal patterns of co-ascription. If it is a true generalization that some dogs are hairless, it is false that no dogs are. [Dog::Not-hairless], as either an entailment or an inference, is incorrect given the existence of even a single hairless dog.

A fourth category of word-pairs is shown in Figure 5.

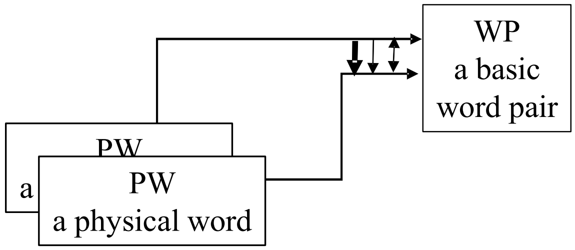

Figure 5. A Word-Pair in a Basic Statement.

This Figure differs from the previous three in showing all three of the connections between two words – entailments, inferences and generalizations. These basic statements were not included in the cognitive psychology experimental paradigm described in Part 1 because, unlike those statements, basic statements provide information about specific objects, and conform to patterns governing both the ascription and co-ascription of concepts. The experimental paradigm in Part 1 is about isolating and measuring each of those two constraints.

Basic statements are constrained by word-to-word associations and, when the words in a word-pair are embodied concepts, also by word-to-object associations. These latter associations express constraints of reference on the statements we make. For any successful counterexample to an entailment or an inference, or any noted exception to a generalization, embodied concepts must satisfy extensional/referential constraints as well as intensional/meaning constraints. If a proposed counterexample to the inference "All cardiologists are doctors" is proffered, then the referent object must be a cardiologist but not a doctor, as agreed to by the relevant linguistic communities. Meaningful true statements must satisfy co-ascriptional constraints; and if their constituent concepts are embodied concepts, they must also correctly name the kind of what they each refer to.

However, this brings to the fore a connection between (i) word-pairs and word-object pairs, and (ii) the kinds of statements we make. Enough may have already been said to make this connection intuitively clear, but I will have more to say about it shortly. For now, I turn from meaning, as expressed in ordered concept-to-concept pairs, and our knowledge of the world, as expressed in empirical generalizations, to reference, as expressed in concept-to-object pairs. This is turning from the topic of concept co-ascription to the topic of concept ascription.

Word-to-Object Connections

A great deal of work in neuroscience has been done on understanding how different images of objects are consolidated in the brain so as to make possible the awareness of an object and, moreover, an object which falls into a specific category. This is a consolidation across episodes of encountering the same object, different perceptual perspectives of the same object, and the apprehensions of the same object in different modalities of sensory awareness.

A monkey will run up a tree when it encounters a tiger, but will leap to another tree when it encounters a snake. So the monkey's non-linguistically-mediated recognition of referents is not just a recognition, on an occasion, of a particular object. It is also the recognition of that object as being an object of a particular kind. A monkey can recognize a tiger as not just a physical object encountered on a specific occasion, but as a *tiger*, as something to run up a tree to escape from. The monkey will run up a tree to escape a tiger, but will take other action to escape a snake. It "knows" what a tiger can and cannot do when encountering trees, and the different things a snake can and cannot do. It also has different warning calls for tigers and for snakes. Those calls are semiotic *signs*, but they are not semantic *concepts,* because nothing but the perceived present image determines whether or not the call is appropriate.[11]

Figure 6 shows what a monkey can do that might look like fully linguistic behavior. It can associate a sign with an object, in the process identifying that object as falling into a specific category of objects. Nonetheless, those signs are not concepts.

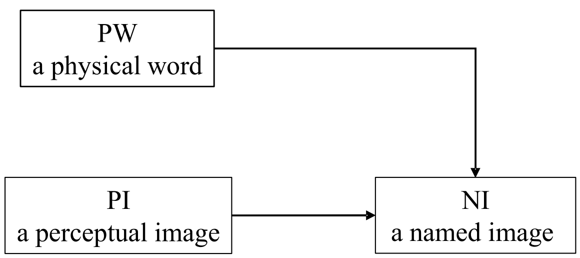

Figure 6. A Named Image.

A monkey's call is a semiotic sign because it is a sound that is associated with a category, e.g. the category of tigers. But the word-to-image association is not constrained by any associations of that sign with other signs. So the result is a named image, but not one named with a concept. Monkeys don't co-ascribe their signs. But these images are nonetheless images of kinds of referents, and in this they are similar to the lexicalized images we language-users have. Recognizing these images is what neuroscientists generally call object recognition, to distinguish it from images that occur earlier in sensory processing streams. Convolutional artificial neural networks (cANNs) have gotten quite good at this.

It is in the animal kingdom that we see referential connections in their pure form. Monkeys have no concepts, no words enmeshed in a network of word-pairs, to associate with their encounters with tigers and snakes. But they know the difference between them. The monkey possesses a LTM image of tigers and another one of snakes, each generalized and consolidated across modalities of sensory perception, different locations and sizes as two-dimensional images on the retina, spatial relationships with other objects in the visual field, perspectival points of view, and so forth.

With monkeys possessing these cognitive abilities, if concepts were no more than what

[11] For well-informed speculation on the nature of animal calls, and whether or not they can be considered elements of a proto-language from which human language developed, see the two books by Derek Bickerton listed in the bibliography. (He thinks they are not elements of a proto-language. I think they are.)

neuroscientists say they are, it would be difficult for them to claim that monkeys don't have concepts of tigers and snakes, or difficult to distinguish between some sense or other in which they do have those concepts, and the full-fledged sense in which language-using human beings have those concepts. Perhaps such difficulties are what have led other researchers to consider the physical words (distinctive sounds) used by monkeys as “proto-concepts”. However, to call them proto-concepts is not to explain *how* they differ from the full-fledged concepts found only in human language. nTM does explain that.

What monkeys and human beings share is object recognition. In the visual modality, these are the most complex representations of the visual stimuli caused by the objects and features we see around us, representations at the end of the visual processing stream which begins in the occipital area, and progresses through the V1-V5 anterior areas to the temporal and frontal areas of the brain. nTM has nothing to contribute to an understanding of how these images are formed, stored and accessed in brains, but it does explain the difference between monkey recognition of objects and the linguistically-mediated awareness of those things. Lexicalized images associate sensory perceptions with embodied concepts, in the structuralist sense of “concept” already presented. Via those associations, the categories corresponding to those additional concepts can be brought to bear on the awareness of the objects. Lacking those associations, they cannot be.

The difference between non-linguistic and linguistic awareness of objects is shown by the contrast between Figures 6 and 7.

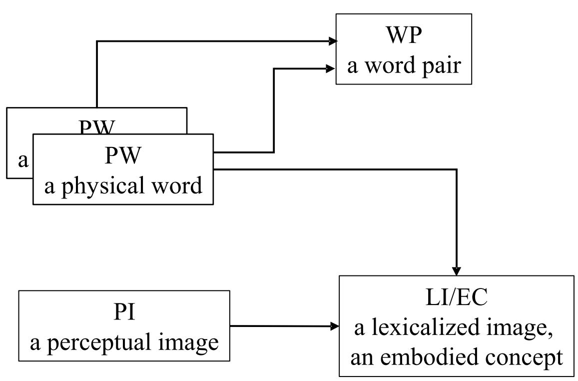

Figure 7. A Lexicalized Image / Embodied Concept.

In this Figure, the PW/WP links are shown as being of equal weight, a diagrammatic device I have thus far used to distinguish generalizations from entailments and inferences. Here the device is used in its neutral sense as representing any of those three kinds of word-pairs. For this same neutrality reason, this diagram does not have any short arrows, since their use is to distinguish those three kinds of word-pairs.

The difference, of course, is that monkeys can name kinds of things with *signs*, whereas human beings can name kinds of things with *concepts*. In Figure 7, and earlier figures, what turns a word into a concept is its co-ascriptional associations with other words.

Some of those concepts are what neuroscientists call concrete words, or else concrete nouns. I call them lexicalized images, or else embodied concepts.

Of course, just as what I am calling word-pairs are, in the brain, pairs of the neural representations of words, what I am calling here word-object pairs are, in the brain, pairs consisting of the neural representation of an object and the neural representation of a word. However, for the most part, I will follow neuroscience literature and omit that terminological interpolation by, e.g. referring to word-pairs rather than "neural representations of word-pairs". On the other hand, I will not refer to abstract representations of concepts in the brain. An abstract neural object is a contradiction in terms.

> Aside. This seemingly minor terminological point is, in fact, a source of considerable confusion in the neuroscience literature. For example, a neuroscientist might reply to my point about the phrase "abstract representations of concepts in the brain" that a better phrase would be "representations, in the brain, of abstract concepts", and that the point of speaking about "abstract concepts" rather than just "concepts" is simply to emphasize that concepts are semantic objects, neither just neurons in the brain nor just phonetic and orthographic sequences outside the brain.
>
> So it would seem that the "rift" that Willem Levelt should have been referring to is the rift between semantic concepts and phonological wordforms. This is pretty much the rift that Descartes first brought into sharp focus and that philosophers have been discussing, as the "mind-body problem", ever since.[12] This distinction between the mental and the physical was brought to the attention of neuroscientists by the philosopher David Chalmers who argued that "the hard problem of consciousness" was how to relate the non-physical phenomenon of consciousness to what goes on in the brain when we are conscious.
>
> On that latter point, I pretty much agree with Stanislas Dehaene (2014, especially pp. 261-262) that we can speak about mental things without giving them equal ontological status with physical things. The impetus to give them equal ontological status, as I argued in Part 1 (especially footnote 20), comes from mixing two distinct hermeneutic contexts of discussion. Just as Dehaene explains consciousness as a global state of the brain, I explain semantics as word-word and word-object learned associations perceived, recorded, and accessed in the brain. In which case, *there is no rift to cross*. In that case, there is also no need for (or possibility of) a location in the brain for abstract semantic objects. And in that case, lacking other plausible candidates, concepts' physical embodiments in the brain can be straightforwardly viewed as the neural representations of the

[12] Levelt's Lemma Model is reviewed in a later section.

orthographic and phonetic sequences which are the words we produce and comprehend. And, as I have been arguing, these groups of neurons become concepts by being subject to a structuralist network which imposes constraints on their co-ascriptions in statements.

The Second Structural Level of nTM

At the second structural level of nTM, word-to-word and word-to-image connections are gathered together into collections which I have called graded sets. None of the figures so far have shown sets of word-to-word pairs or sets of word-to-object pairs, although I have alluded to such sets already, and I described their structures and processes at the level of analytical philosophy of language and lexical semantics in Part 1.

These sets are graded according to the strength of their word-to-object or word-to-word connections, as measured by ascriptional and co-ascriptional patterns, as described above and in Part 1. Ascriptional connections relate a word to a perceivable object; co-ascriptional connections relate a word to another word. Each play their role in basic statements which involve embodied concepts. Word-to-word connections play their role in all statements, no matter how abstract their referent objects and features may be.

Figure 7 is also a diagram of how semantics – comprised of meaning, reference, and their mutual influences – is realized in the brain. All concepts exist as wordforms whose learned co-ascriptional links to other wordforms, in word-pairs, constrain their use in the statements we make and comprehend. Embodied concepts also have links to images, and to judge that sometimes a word does not refer to the object represented by an image is to judge that a mistake in the ascription of a word to an object has been made. I can't call an object speeding along an expressway on four tires a horse, because anyone can see that it is not a horse. Similarly, a monkey would make a mistake by uttering the sound for tigers when snakes were in the vicinity, but not tigers.

So, for each concept, we have its physical wordforms (phonetic and orthographic) as the anchors of the brain's representation of it. Entailment- and inference- strength co-ascriptional associations with other wordforms establish an ordered activation relationship from the concept to its related concepts, each wordform being a concept because of its place in a network of mutually-constraining wordforms. The graded strength of each concept-concept relationship manifests itself in the degree of our willingness to require that a second concept be ascribable to a referent to which we have already ascribed a first concept. The co-ascription is manifested in basic statements – either ones directly produced, or the basic statements constituting a decomposition of more syntactically-complex statements. And there is a reciprocal set of "anti-related" concepts in which applying a concept leads to a disposition to disallow the ascription of another concept to the same referent, this too being a graded matter-of-degree phenomenon.[13]

[13] For convenience, I will use the term "Graded Meaning Set" to include these anti-related concepts with the related ones since, clearly, both are determinants governing ascribing, or withholding the ascription of, a concept, given another concept having already been ascribed to that same referent. I also point out that while I focus on co-ascriptions within a single statement, co-ascriptions also apply across statements, and even across speakers in the same conversation.

As our consideration of monkeys has shown, even without language we can have neural representations not just of particular objects and features, but also of *kinds* of objects and features. These pairings of words and objects are the way we (and the monkeys) use words to refer to objects, and to refer to them as instances of categories of objects. However, for monkeys, there are no Graded Meaning Sets involved, and so there are no word-pair constraints on the use of signs. Lacking such constraints, monkeys cannot draw conclusions about tigers that rely on categories associated with the tiger category such as, for example, hunts in the early evening. Monkeys have no inter-category connections.

But with language, things are different. A Graded Meaning Set for a given word can be associated with a Graded Meaning Set for a related word, since Hebbian learning has made it possible for an encounter with a referent to trigger activation of the neural wordform saying what kind of thing the referent is, and because that wordform is enmeshed in a network of co-ascriptional constraints. This network requires assent to the co-ascription of certain second concepts, permits the co-ascriptions of other second concepts, and disallows co-ascriptions with yet other second concepts. Correlatively, activation of the wordform of an embodied concept can trigger activation of the memory of a (specific or undetermined) member of the category which that concept names. The objects which fall under that embodied concept category constitute the Graded Reference Set for that lexicalized image. They are the members of that set.

Of course, just as with meaning sets, membership in a reference set is, as all neurally-based representations are, a matter of degree. For example, there is the well-known series of drawings, presented left-to-right and all at once, in which the left-hand drawing is clearly a drawing of a cup, and the right-hand drawing is clearly a drawing of a mug, and in which each drawing is very similar to its nearest-neighbor drawings. But towards the middle of the series, if asked "Is that a drawing of a cup or a mug", it becomes difficult to say which.

There are also prototypes to consider. Descriptions of the prototype effects of concept reference are commonplace. However, "prototype" suggests that all such graded referential associations of a word with a physical object or feature, cluster around a "referential nucleus" which picks out a notable subset within the set of all valid references.

This hypothesis, however, is mistaken. The reason we respond with "fruit" to images of apples more rapidly than to images of cranberries is not that apples are more prototypically fruit than cranberries are (whatever that means). [Apple<--->[Fruit]] is a stronger referential association than is [Cranberry<--->[Fruit]] for some such reason as that [[Fruit]<--->[Sweet]] is an established generalization, and although both [[Apple]--->[Fruit]] and [[Cranberry]--->[Fruit]] are entailments, [[Apple]<--->[Sweet]] is a stronger generalization than is [[Cranberry]<--->[Sweet]]. This is another example of how reference does not exist, when carried out with concepts, in isolation from the meaning of those concepts.

This shows that many Graded Reference Sets require no prototype at all at their centers. In a gradient of colors, arranged from red to blue, there is no prototype color. In the set of NFL football teams, there is no prototypical team (in spite of the Dallas Cowboys claiming to be "America's Team"). In number theory, there is no prototypical number.

> Aside: I here define the expression "neural group" to refer to the association of a localized group of neurons, together with specific patterns of activity, with an identifiable cognitive function. This means that I consider two distinct patterns of firing, within essentially the same set of neurons, as possibily distinct neural groups. The identifiable cognitive functions focused on in this essay, of course, are linguistic ones.
>
> There are neural groups corresponding to major functions such as object recognition, or executive management. But I also count as neural groups those corresponding to specific concepts, and/or to specific images, which are accessed in the processes of comprehending and producing both individual lexical items, and the statements which combine them.
>
> Some cognitive functions – those specified at a high level, such as awareness or consciousness – will result from the sequential or concurrent activity of many neural groups. But until such holistic functions are broken down into their localized components, they are not understood. To say that a function X happens everywhere in the brain, or that a unit of information is stored everywhere in the brain, is to say nothing useful at all. At a high-enough level, it is a truism. At that same level, it is uninformative.

I begin with Graded Meaning Sets.

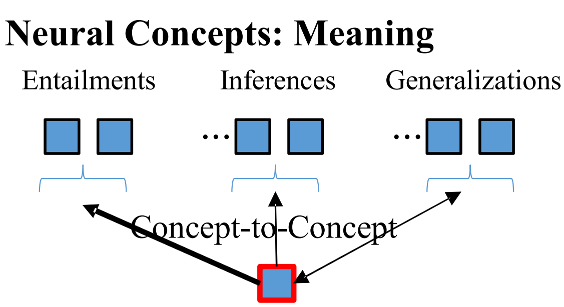

Figure 8. The Graded Meaning Set for a Neural Concept.

> *As in Figures 2 – 5, the heavy and light uni-directional arrows indicate, respectively, entailment and inference connections, and the bi-directional arrow generalizations. Each box is a neural concept, whose physical form is a group of neurons that fire, or fire in a specific pattern, when a lexical item is heard or read. The red-outlined box is the concept whose intensional constraints*

are implemented in these relationships. The arrows represent dispositions for one of the related concepts to be used in a co-ascriptional pair with the first concept.

Aside: note that uni-directional arrows can be traversed in the opposite direction. If "All X are Y" is true, and Xs exist, then it follows that "Some X are Y" is true, and from that it follows that "Some Y are X" is also true. This is, in fact, the distinction between the Aristotelian interpretation of universal quantification, in which Xs are assumed to exist, from the Fregean and still-current interpretation in which they are not assumed to exist. The brain, I propose, prefers the Aristotelian interpretation.

In both our commonsense understanding of meaning, and more formal contexts, we say that the meaning of a concept is expressed in its (formal or informal) definition. One representation of definitions is as a boolean combination of entailed concepts – as in the compositional semantics definitions of the semantic theory Fodor and Katz provided for transformational-generative grammar. More broadly speaking, the meaning of a concept is the complete set of its entailment- and inference-related concepts, and also those concepts related to it in (existentially-quantified) generalizations.

Each concept in Figure 8, of course, is itself a red-box concept.[14] This makes the dispositions to require, permit, or reject the co-ascription of paired concepts a set of constraints which gives each red-box concept its meaning. These dispositions are tendencies to fire or not fire after a red-box concept has fired as the first concept in a basic statement.

Each concept in Figure 8 is also part of many other concepts. [Adult], for example, is part of the meaning of both [Man] and [Woman]. This intricate network of transitive constraints ties together auditory wordforms in one neural lexicon, and orthographic wordforms in another neural lexicon. Via such processes as silent reading, in which some of the words we read are silently sounded out, the two lexicons, for each of us and at each point in time, are kept roughly synchronized. I will have more to say about this in the following sections.

Neural Concepts: Reference

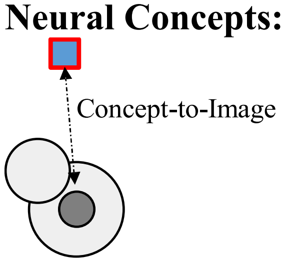

Figure 9. The Graded Reference Set for a Neural Concept.

[14] For those with black-and-white copies of this essay, a red-box concept can be identified as a concept related to an image, and/or a concept for which entailment and inference arrows point outwards.

This Figure shows a concept bi-directionally related to a graded set of images. These images exist in long-term memory, and on any encounter with a previously-encountered object, the short-term memory awareness is associated with and also modifies the long-term image it most closely resembles.

Figure 9 shows a core set of LTM lexicalized images of an object, a dog say. The set is surrounded by less immediately recognizable images of that object, and the figure also shows that lexicalized images can overlap. When a currently perceived image is more closely matched to this consolidated image than to others, the neural wordform “Dog” is activated. Seeing a dog results in my readiness to say something in which I will use the word “dog”. Conversely, hearing someone telling me to notice something about a nearby dog, the neural wordform “Dog”, together with an indexical expression or gesture, enables me to pick out the alluded-to dog from all the other objects in the current scene.

What analytic philosophers continually emphasize, when they are emphasizing semantic fundamentals, is that language is *about* things. If I say, of a cat that has just jumped up on the kitchen counter, “The cat is on the mat”, I'm wrong, and I'm wrong because that isn't how things are. The statement doesn't correct describe what it is about. The cat is on the counter, not on the mat. Any semantic theory – such as a pure form of Saussurean structuralism – that has nothing to say about the aboutness of language, is either simply wrong, or at least seriously incomplete. My view is that it is incomplete. The conceptual network of meaning is a structuralist network, but it needs to be combined with a theory of how concepts refer to things and how they describe them, and how those things make our statements about them either true or false. nTM does this.

As we encounter a repeated pattern in which a word is used to designate the kind of an object, we come to use the word in that way ourselves. The fuzzy edges of most lexicalized images are due to variations over multiple occasions of word-to-object ascription which reflect the shifting predilections of different people at different times to include or exclude a presently encountered referent in the range of one or another kind. The referential use of the word has a widely-accepted core range, whose wide acceptance is both the cause and the effect of its frequent pairing with a perceived image, and its infrequent rejection as a response to a recognized perceptual pattern. Other acceptable referential uses often exist as a penumbra surrounding core usages, shading even further into uses which may be tolerated as occasional extensions of referential range, but not as regular ones.

The processes by which individual perceptual images contribute, in LTM, to a consolidation of those images, are beginning to be understood by neuroscientists. The hippocampus is centrally involved in the consolidation process, but the consolidated image itself, continually modified by those hippocampal processes, can subsequently be accessed without involvement of the hippocampus. We can remember what things generally look like without recalling specific episodes involving those things.

Neural Concepts and Neural Semantic Networks

A single concept in a neural semantic network is shown in Figure 10. Hereafter, I will often refer to concepts as neural-lexical-units, or NLUs.

> Aside: sometimes I will speak of NLUs as concepts, with the qualifier “in the brain” elided. This is usually when the context is closely focused on the neurophysiology. At other times, I will speak of NLUs as *representing* concepts. This is usually when the context is focused more on the analytical philosophy of language and lexical semantics perspectives on meaning and reference. The latter way of speaking brings a mental/physical dualism to the forefront; the former way of speaking reflects my own semantic reductionism. I am content to develop this semantic theory in terms of this kind of reductionism, and leave the ontological issue of physicalistic monism vs. mind-body dualism to the philosophers who study such things. It's all a matter of what kind of progress one wants to make, and in what hermeneutic context.

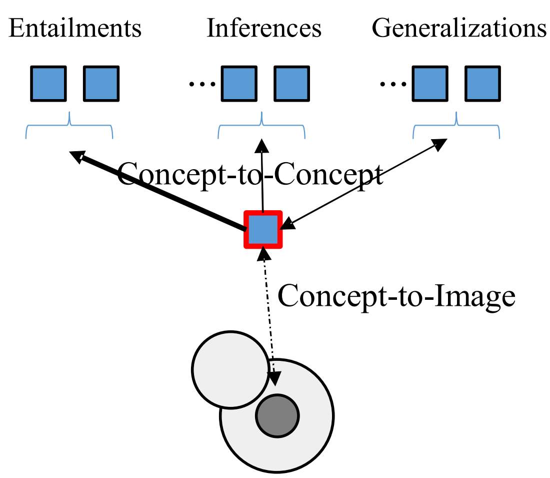

Figure 10. A Neural Concept: Meaning and Reference.

> *As shown, neural concepts (NLUs) have relationships to other concepts, which I have distinguished as entailment relationships, inference relationships and generalization relationships. Concepts which refer to perceivable things also have relationships to integrated images of those things. These latter concepts I call embodied concepts, and neuroscientists generally call them concrete words. Concepts which are not embodied I call abstract concepts. As explained above, and below, all these relationships are graded, i.e. they exist on a continuum.*

Suppose that the object I have just noticed, which was pointed out as being a dog, turns out to be some hyper-realistic Japanese robot. Since [Dog::Mammal] and [Mammal::Warm-blooded] are both entailments, and robots are neither mammals nor warm-blooded, I am forced to admit that what looks exactly like a dog is not, in fact, a dog. Meaning has trumped visually-based reference. This, along with numerous examples in Part 1, illustrate that reference by means of concepts must be consistent with the entailments and inferences of the referring concept. The red-box concept in Figure 10, the one whose definition is the combination of its entailment- and also inference-related concepts – is subject in its ascriptive uses to the co-ascriptive constraints of its Graded Meaning Set, and is subject in its co-ascriptive uses to the ascriptive constraints of its Graded Reference Set.

But individual NLUs are more intricately knit in their semantic networks than may be apparent. So I attempt to show what that semantic network is like, in Figure 11.

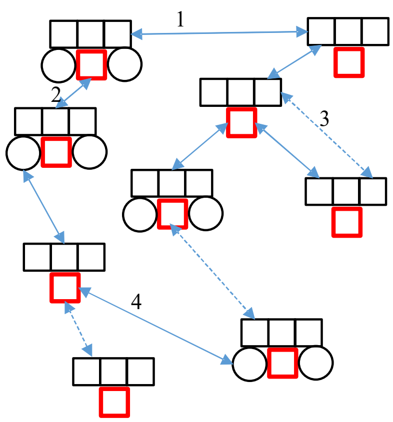

Figure 11. The Neural Semantic Network.

This Figure uses a condensed icon to represent individual NLUs; each one is a condensed form of what is shown in Figure 10. For embodied concepts, the icon includes circles representing the images which their associated red-box concepts lexicalize. The red box represents the concept which names that image, e.g. as an integrated image of the kind [Dog]. For both embodied and abstract concepts, the non-red-boxes represent the related concepts in the red-box concept's Graded Meaning Set. Box-to-box solid-line arrows do not represent relationships; they indicate that the two boxes are the same concept, encoded as the same neural wordform. Box-to-box dotted-line arrows represent antonymous connections, e.g. [Animate] and [Inanimate]. Box-to-circle arrows represent alternative lexicalizations of the circle-represented images, e.g. that [Animal] could be used instead of [Dog] on a given reference occasion.

The two boxes on Arrow1 might represent the concept [Adult]. One might be entailed by the red-box concept [Man] and the other by the red-box concept [Woman].

The two boxes on Arrow2 might represent the concept [Man]. At the upper end, the concept receives its meaning from its related second concepts. At the lower end, the concept contributes its meaning to the meaning of that lower-end red-box concept.

Arrow3 might relate the antonymous concepts [Man] and [Woman].

Arrow4 represents an alternative lexicalization of an image. The red-box on the NLU without circles might represent an abstract concept like [Animate physical object] while the red-box on the NLU with circles might represent an embodied concept like [Streptococcus bacterium]. In some contexts, [Animate physical object] could be used (with an indexical) to direct attention to a streptococcus bacterium.

> Aside: when a neuron fires, that spike represents the summation, at a point in time, at the axon hillock, of the positive and negative inputs to the dendrites and soma of that firing neuron. And so when that spike makes its contribution to the fire-or-don't-fire decision of downstream neurons, the information conveyed by that spike to those neurons is that a combination of positive and negative inputs to its own neuron caused it, at an earlier point in time, to fire.
>
> Now note this analogy, that in the description of Arrow2, the red-box concept contributes its own semantic “summations” to another concept in which it occurs as a non-red-box concept, i.e. as a second concept in a co-ascribable pair. Meaning spreads through the semantic web just as neural signals spread through the neural web.
>
> There is a difference in time-frames, of course. Neurons change their relevant states from millisecond to millisecond; concepts change their semantic states, in each human brain and even more so across societies of language users, more slowly. Dictionaries, for example, are revised on a time-scale of something like decades. Useful insights might come from pursuing this analogy to the point where it is more than an analogy, but I won't say anything more about that topic in this essay.

The remaining arrows will be self-explanatory, and are there only for illustrative purposes. However, one misleading thing about Figure 11 is that it makes it appear that concept-to-concept pairs are relatively sparse in the lexicons. In fact, of course, they are not. A robust concept might have a dozen or more entailments, many dozen inferences, and many more generalizations, and an embodied concept might have any number of overlapping images it is associated with. Every concept is a red-box concept, with its own often numerous connections with its second concepts and with lexicalized images. And, as should by now be apparent, concept-to-concept connections are semantically transitive (just as neural signals are electrochemically transitive). If [A::B] and [B::C] are entailments, for example, then [A::C] is, too. If either or both are inferences, then [A::C] is also an inference.

> Aside: I note here, and will discuss later, that when we are aware of the meaning of any concept without, as is always the case, being explicitly aware of each and every one of its concept-to-concept connections, i.e. of each component of its definition, what we are aware of is a *gestaltist immediacy* of these related concepts and the strengths of their connections, out along a semantically transitive series of links that, ultimately, relates each single concept to every other concept. As they exist in LTM, I will sometimes call these NLUs, which (except when being first learned) enter awareness as gestalts and are recorded as *engrams*.

Of all these related concepts, however, the ones most semantically dominant in our awareness will be the related concepts that have already appeared in the conversations we are involved in, or the text we are reading. I will discuss this point later when I describe the processes which give rise to producing and comprehending coherent narratives.

Concepts and Statements

Concepts come together in statements, which are the units of language in which information is expressed and comprehended. The neurophysiological basis for statements can best be described by distinguishing several categories of statements, organized along a continuum.

Logicians distinguish three kinds of statements. Each has a distinct role to play in this semantic theory, and given that language is a form of patterned human behavior, these roles must be explained in terms of neurally-based dispositions manifested in that behavior.

Entailments and Inferences

One kind of statement is a *universally quantified statement*, such as "All A are B". This statement says that if anything is [A], then it is also [B]. In nTM, it is neurally represented as the ordered concept pair [A::B]. If we assent to this statement, then whenever we are willing to agree that something is [A], we will be willing to agree that it is also [B]. A neural group implementing this concept pair will manifest as a pattern of linguistic behavior in which we will not propose or accept as true any basic statement ascribing both [A] and [not-B] to whatever thing that statement refers to.

Some universally quantified statements are true "no matter what", true in any and all imaginable circumstances. These are called *a priori* statements by philosophers. Others happen to be true, that is, they are true because of what the world is in fact like. These are called *a posteriori* statements. A closely-related distinction is between *analytic* and *synthetic* statements. Analytic statements are ones true solely because of the meaning of their component concepts; synthetic statements are ones whose truth also depends on their correctly describing what they are about. These related distinctions were discussed in Part 1, where the vexing issue of whether any statements can be both synthetic and a priori, or analytic and a posteriori, was resolved.

On this topic, the conclusion of Part 1 was a Quinean one. Analytic/synthetic, and a priori / a posteriori, are continua, not dichotomies. Nonetheless, there are statements whose denial would be almost always, by almost everyone, considered a semantic mistake, and I will find it use to retain the labels "analytic" and "a priori" for them.

The point to be noted, for this account of semantics, is that a compositional semantics account of the meaning of a lexical item is given by the set of analytic statements in which the concept being defined is the first one, e.g. that the definition of "bachelor" is given by the set "All bachelors are men" and "All bachelors are unmarried", or that the definition of "men" is given by the set "All men are human beings", "All men are male" and "All men are adults".

It is no coincidence that the principal philosopher to develop an early theory of compositional semantics – which TM uses in explaining the meaning of words – is also the philosopher who proposed the Amodal Symbolic Model called the Language of Thought (LOT). That philosopher, of course, was Jerry Fodor. From this perspective, TM can be seen as a theory which modifies the conceptual atomism of LOT by incorporating those atoms in a structuralist network within which Fodor's own compositional semantics account of definition (but with the exclusion of the notion of semantic primitives) fits comfortably. Nonetheless, TM still explicitly rejects any implicit or explicit Fodorian notion of concepts as being abstract objects in the brain.

Equivalently, the meaning of [Bachelor] is given by the set of ordered concept pairs [Bachelor::Man], [Bachelor::Unmarried], and the meaning of [Man] is given by [Man::Human Being], [Man::Male], and [Man::Adult]. Hebbian associations among these ordered concept pairs should be discoverable by current neuroscience experimental technology, which would then provide a neuroscience account of the person- and occasion-relative meaning of those concepts. I do not believe that these experiments have yet been carried out in current experimental cognitive neuroscience.[15] The reason, of course, is clear; currently, neuroscientists are not aware of the structuralist web of meaning, and continue to think of concepts as discrete nodes in an abstract conceptual lexicon located in some as yet unidentified area of the brain.

Another task is a sociolinguistic one of explaining how the meaning of those concepts – as types, not as person- and time-relative tokens – is abstracted by lexicographers as they compile and revise their dictionaries. That task was completed in Part 1.

The set of analytic statements in which a concept is the first concept is one group of ordered concept pairs on the statement continuum. Next to it is a group of ordered concept pairs which also express universal quantifications – ones of the “All A are B” form – but ones which are not analytic. So we may distinguish them as synthetic statements, keeping the matter-of-degree hypothesis in mind. Denials of synthetic universal quantifications are not considered semantic mistakes; they are just mistakes about what happens to be true.

Each of this second category of universally-quantified statements is an *inference* from the first concept to the second concept, just each analytic universally-quantified statement is an *entailment* – a relationship true in all possible worlds, not just in this one. These entailments and inferences, for a given concept, constitute the Graded Meaning Set for that concept. These Graded Meaning Sets are embedded in a structuralist network of, ultimately, all concepts because the second concepts in those entailments are also first

[15] Neuroscience word association experiments do not fit the bill because they do not distinguish between ordered and unordered word-pairs. Another reason they do not fit the bill is that they do not distinguish between entailments and inferences.

concepts in their own Graded Meaning Sets, both transitively and recursively.

These sets are graded sets because their pair relationships are graded. Entailment is a matter of the inconceivability of exceptions, and there is no sharp line distinguishing the conceivable from the inconceivable.[16] This is true of universally-quantified statements (ordered concept pairs) at the level of statement types. But it is more fundamentally true at the level of statement tokens, where what is inconceivable to one person or group of persons, at the same or different periods of time, will differ across those persons and times.

> Two of the jobs for a neuroscience semantic theory are to (i) describe the neural wordform pairs corresponding to the entailments in Graded Meaning Sets, and also (ii) describe the pairs corresponding to the inferences in Graded Meaning Sets, and then look for them. I have described them in this essay. It is now up to experimental neuroscience to look for them. The cognitive psychology experimental paradigm I developed in Part 1 should be a good guide. I also indicate there the use to which that data could be put by lexical semanticists and lexicographers.

Generalizations

A third kind of statement is an *existentially quantified statement*, such as "Some A are B". This statement can be expressed as the unordered concept pair [A:::B] – unordered because if some [A] are [B], then some [B] are [A].[17] An NLU implementing this concept pair will manifest as a pattern of linguistic behavior in which we may or may not accept as true a statement that a particular thing which is an [A] is also a [B] (or vice versa). These unordered concept pairs are a third category of concept pairs along the Graded Meaning Set continuum being described here – entailments and inferences being the first two. They belong on this continuum because they are concept pairs, and like the other ones, their links express dispositions to require, permit or forbid describing something as instantiating one but not both of those concepts. All of these concept-to-concept dispositions are sets of co-ascriptional links which constrain the use of pairs of concepts in basic statements.

It might seem that existentially-quantified statements provide no semantic constraints because they permit the co-ascription of [A] and [B], but neither require nor forbid it. But this is wrong. From “Some A are B”, we cannot conclude that “All A are B” is either true or false. But we can conclude that the negation of “All A are B” is false – that it is false that no [A] are [B]. In logical notation, while we can conclude nothing about $\forall x(Ax \supset Bx)$, we can conclude that $\forall x{\sim}(Ax \supset Bx)$ is false. In this way, through constraints on the

[16] See Part 1, pp. 23-25, and the references there to Lyons, 1995.

[17] Note that I use the notation [A::B] to represent ordered concept pairs, and [A:::B] to represent unordered concept pairs.

possibility of accepted counterexamples – which is the paradigm form, in ordinary language, in which meaning and reference putatively conflict – existentially-quantified statements contribute to those constraints as much as universally-quantified ones do.

Inferences are a bridge between entailments and generalizations. They are the "All A are B" statements which could be denied without misusing language, without making a semantic mistake (as judged by an interlocutor, or by a community of speakers). They are statements to which we can imagine counterexamples although we don't believe that there actually are any. But it's not much of a step from there to a belief that there might be in fact a few counterexamples; and indeed, from one occasion to another, any one of us might waver back and forth between "no actual counterexamples, albeit conceivable ones" and "perhaps one or two counterexamples".

Existentially-quantified statements express incomplete information. If we say that some [A] are [B], this fails to distinguish, among the things that are [A], those that are also [B] from those that are not. But incomplete information like this can still be useful, indeed essential. If we believe that most [A] are [B], and that things that are [B] are dangerous things, then we're going to be careful in approaching and handling things that are [A]. If we believe that only a very few [A] are [B], and that things that are [B] are unfortunate to encounter but not deadly, we may not be as careful.

I call existentially-quantified statements *generalizations*. Like inferences, generalizations are also graded, from those we think apply in nearly all cases to those we merely suspect may be true. Generalizations which lie along a continuum from strong to weak, are followed by generalizations we are completely agnostic about, and from there to those we suspect might be false, to those we think probably are, onto those we are almost sure definitely are.

> A third job for a neuroscience semantic theory is to describe the neural wordform pairs corresponding to Graded Generalization Sets, and then look for them.[18] I have described them in this essay. It is now up to experimental neuroscience to

[18] Generalization sets were called association sets in Part 1. A more serious source of terminological confusion between Parts 1 and 2 is that in Part 1, I dropped the term "entailment" and used the term "inference" for all universally-quantified statements. Here in Part 2, I keep the term "entailment" for analytic statements, with the Quinean proviso that no analytic statement is completely immune to revision if patterns of co-ascription change widely enough and for long enough.

And one more bit of terminology. In draft form, I found myself frequently using the awkward phrase "entailment-, inference- or generalization-strength connections", or "entailments, inferences or generalizations". These are the three categories of word-to-word connections that make up the meaning, the Graded Meaning Sets, of concepts. I have now revised the text to often use the adjective "intensional", as in "intensional relationships" instead. Justification for this terminology, in the form of the philosophical distinction between the *intension* and *extension* of a concept (in Fregean terms, between its *sense* and its *reference*), can be found in Part 1. I have adopted these terminological changes because I think that, on balance, they do more good than harm.

look for them. The cognitive psychology experimental paradigm I developed in Part 1 should be a good guide. I also indicate there the use to which that data could be put by lexical semanticists and lexicographers.

Neural/Semantic Compositionality

Linguists have pointed out, at least for decades, that the meaning of statements is compositional. It is a function of the meaning of its component concepts. TM explains this in its analysis of basic statements.

Basic statements are a *fourth* kind of statement. Unlike the other three, basic statements contain no variables; they are *fully instantiated statements* such as “That A is B". This statement identifies a particular object as being both an instance of [A] and of [B]. A statement containing this concept pair will lead to a pattern of linguistic behavior in which we will accept the statement as true only if we agree that the designated object is both an [A] and a [B]. As already explained, this requires that the statement satisfies referential constraints, that [A] and [B] each correctly name the kind of thing or feature they each refer to. It also requires that the statement satisfies meaning constraints, that the co-ascription of [A] and [B] produces no contradictions across the set of concept-pairs in the union of the Graded Meaning Sets of both concepts.

As described in Part 1, [A] will indicate what kind of thing the object is, while [B] will indicate some feature it is said to possess. [A] and [B] are both concepts, and the distinction between concepts indicating what kind of thing a designated object is and concepts indicating some property or relationship the designated object has, is not a distinction between two kinds of concepts. Rather, it is a distinction between different roles a concept can play, one role being that of helping to pick out some thing, and the other role being that of saying something about what has been picked out.

In basic statements, one concept, together with an indexical, constitutes the subject term of the statement. Its role is to pick out the referent of the statement by indicating what kind of thing it is, such as a horse, i.e. an instance of the concept [Horse]. A second concept then is used as the predicate term of the statement. Its role is to say something about the referent, such as that it is located in a pasture. What the statement is about instantiates, not just [Horse], but also [Pasture] (or [Located in a pasture]).

This is illustrated in Figure 12.

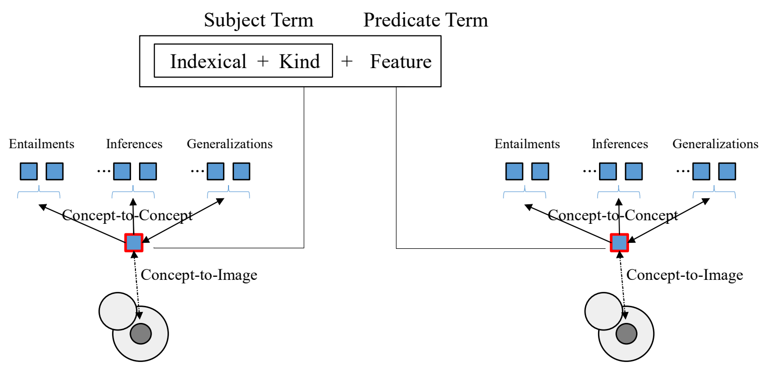

Figure 12. A Basic Statement.

The red-box concepts represent the two concepts co-ascribed in a basic statement, one picking some thing out with the help of an indexical, and the other picking something out that co-occurs with the first object, either as a feature or perhaps as a role in a process. The reference sets must both be satisfied, e.g. we can't say that a pile of sand is a rock; we can just see that it isn't. The meaning sets must not conflict, i.e. the union of the ordered word-pairs in the meaning sets of the two concepts must not result in one or more contradictions.

Fully instantiated statements are the basic statements in which meaning and reference combine to influence one another and, in the process, express information. In the statement "That A is B", some particular thing is picked out by indicating what kind of thing it is, and then using an indexical to point to a particular instance of that kind. That much is the job of the subject term of the statement. The statement then says that the designated thing possesses the property or relationship indicated by the predicate term of the statement, the property or relationship of being a [B].

Here meaning and reference come together. If we don't agree that the designated object is both an [A] and a [B], then agreement in reference has failed. If we agree that it seems to be an [A], but one or both of us believes that it can't be an [A] because all [A]s are [C]s, and that the referenced thing definitely isn't a [C] and so can't be an [A], we have entered the semantic space in which meaning and reference conflict. But if we agree that the evidence of our senses shows that the designated object is both [A] and [B], and we can find no [C] or [C]-like concept whose attribution to the object would indicate that, contrary to appearances, it can't be an [A] (or a [B]), then we have reached agreement – this time, at least. Our referential use of these two concepts are in agreement (as far as we can tell), with the meaning we attach to each concept.

Is the statement "That color is attractive" a basic statement? It would seem that it is not, because it refers to a color, i.e. to a concept, and not to a particular instance of a concept. But on closer analysis, we can see that [Color] is the kind of the subject term's referent, and [Attractive] is an instance of a feature we ascribe to it. [Color] is the kind; as in all basic statements, when combined with an indexical ("That"), it picks out a member of the set designated by the kind. That particular color I am referring to is a member of the set [Color]. Similarly, the set of all attractive things is the kind of which the attractiveness of that particular color is an instance. So in fact, “That color is attractive” is about a specific instance of the kind [Color] and of the feature [Attractive], just as all referents in basic statements are about co-instantiations of their two concepts.

Or consider a more difficult example, “Color makes the world beautiful”. Here we do have a definite counter-example to the basic statement structure in which the subject term consists of a concept and an indexical. There is no indexical here, and that is because the referent is not a particular thing; the category itself is the referent. As I described in Part 1, basic statements whose subject-term referents are concepts/categories, and not instances of them, still pick some thing out, and still say something about it. There in Part 1, I suggested that, in such statements, there is a usually implicit superordinate category under which the referent category is understood to fall. In this example, it might be something like [visually perceivable feature of things].

I note that, mathematically, the difference between this type of basic statement, and the type I have focused on in this essay, is that in most of the basic statements I have discussed, the referent is a member of the set indicated by the subject category, whereas in this example, the referent is itself a set, related to its implicit categorizing set by set inclusion. As explained in Part 1, the MRRT interplay between meaning and reference is an interplay between dynamic set membership forces (expressing the referential range of the set) and dynamic set inclusion forces (expressing the meaning of the set). In the “Color makes the world beautiful” basic statement, the relationship is one of set inclusion. There is no extensional constraint here; all is intension. Such, broadly speaking, is the case in all basic statements both of whose concepts are abstract.

> A fourth job for a neuroscience semantic theory is to describe the neural wordform pairs corresponding to basic statements, and then look for them. I note in particular that this is a more restricted search through linguistic corpora than current word-pair searches which also include non-co-ascribed word-pairs. These latter searches add nothing but noise to the semantic data of co-ascribed word-pairs. I have described these neural wordform pairs in this essay. It is now up to experimental neuroscience to look for them. The cognitive psychology experimental paradigm I developed in Part 1 should be a good guide.

The Third Structural Level of nTM

As thus-far described, nTM could be realized in a network of silicon neurons just as well as in a network of biological ones. I now turn to the embodiment of nTM in human beings. I begin by distinguishing five networks in the language-using brain.

The Modality-Specific Network. One is a distributed network of modality-specific images of physical objects, features and processes. This is the sensory raw material of conscious awareness, although by the time it reaches awareness, it is already extensively processed and integrated. Broadly speaking, each image begins in the primary processing area for that sense input, and then becomes an image available to awareness after being processed in occipital, parietal, posterior temporal and, finally, fronto-temporal and prefrontal areas. For example, consciously-accessible auditory images are located close to Heschl's gyrus in the temporal lobe, and consciously-accessible visual images are located out beyond V5.

The Integrated Image Network. A second network is a meta-network which integrates those modality-specific images into unified representations of physical objects. This network is mainly located in the fusiform gyrus and the inferior temporal lobes for static images of objects, and in posterior parietal areas for dynamic images of objects, connected by the inferior longitudinal fasciculus. It is a *cross-episodic* integration, enabling us to recognize the same objects across many different encounters. It is a *multi-modal* integration, enabling us to recognize the same object by sight, sound and feel (and taste and smell too, if relevant). It is an *omni-perspectival* integration, enabling us to recognize an object as the same object across occasions of seeing it from the front, back, side, top or bottom, and in any spatial orientation relative to the ground. And it is, finally, a *contextualizing* integration, enabling us to locate objects as they move through the environment, and also in their relative locations to other moving objects in that environment. I will call images which are integrated across the four dimensions of episode, sensory modality, perspective, and context, *integrated images*. Integrated images are full-fledged images of objects in the physical world. Note that this network and the first network both exist in the brains of non-language-using animals, as any observation of cats reacting to the presence of dogs will illustrate.

The Physical Wordform Network. A third network is a network of phonetic wordforms and of sound-alike phonological links among them, approximately situated in the left mid-superior temporal lobe, and essentially including the cortex of that area. This network is often called the phonetic network. As a phonetic network, it has no semantic significance by itself. "Looks" and "books" are phonetically similar, but they are not semantically related. There is also a secondary network of orthographic wordforms, primarily located in the left posterior fusiform area (left occipto-temporal sulcus), medial to the inferior temporal gyrus, known as the Visual Wordform Area (VWFA). Again, purely as wordform images, this network has no semantic significance.

The Graded Meaning Set Network. A fourth is a network of concepts and of co-ascriptional links among them, assembled into Graded Meaning Sets. nTM proposes that this network is co-located with the phonetic and orthographic networks. Wordforms are phonetic and graphemic sequences, and in the phonological network, they exist in a sea of other phonetic sequences.[19] These wordforms are NLUs, the neural representations of concepts. The meaning links between ordered pairs of these wordforms are synaptic connections among neural groups – learned associations of learned associations of phoneme/grapheme sequences. nTM places these Graded Meaning Sets in the neural wordform areas – the phonetic wordform area, and the orthographic wordform area. This makes each wordform area the neural representation of a semantic lexicon, each coordinated with the other. There are no abstract concepts, stored anywhere else.

The Graded Reference Set Network. In a fifth network, the integrated images of objects and their features are associated with the wordforms that refer to those objects and features. Since these wordforms are also integrated, in a co-ascriptional network of Graded Meaning Sets, they have semantic content, and thus, when combined with integrated images of objects, they make those images what I call *lexicalized images.* Lexicalized images are the "roots of reference" (to borrow a phrase from Quine). Neuroscience data indicates, however, that lexicalized images are not co-located in the lexicons with the concepts which lexicalize them, and so the nTM Semantic Model (see Figure 13) shows them existing in what I call the embodied concept hub area, where they combine pointers to integrated images with pointers to the wordforms in those lexicons which lexicalize them.

These five networks come together in the Semantics Network to create nTM's Neurosemantic Model.

The Semantics Network. In an integrating network, meaning and reference reciprocally influence one another in our statement-making speech acts, as already described earlier, and also in Part 1. Given that the Meaning Set and Reference Set networks are not co-located, this suggests that the reciprocity of meaning and reference is dependent on the fasciculi that connect those two non-adjacent areas.[20]

[19] I will have more to say about this later. Here, I just note that the sea of all phonetic sequences is a far larger sea than the sea of all phonemic sequences, when "phonemic" is used in the sense of "smallest unit of human speech that has meaning". A phonetic sequence is a sequence of speech sounds, whether or not they constitute a unit of meaning. So a wordform is a phonetic sequence which is also a phonemic sequence. "Phoneme", in this sense, is roughly synonymous with "morpheme"; the only difference is in where the emphasis is placed, on sound or on meaning.

[20] And this has experimental implications. If Meaning/Reference Reciprocity depends on fasciculi connecting these distinct neural areas, then lesions to the fasciculi relating them, but not to either neural area, should result in the following verbal behavior: (i) intact ability to name kinds of things; (ii) intact ability to define those kinds; but (iii) inability to draw conclusions about those being-perceived named things, conclusions which would follow from entailments, inferences or generalizations about them. If

Figure 13 shows the functional areas which I propose as the distinctly localized areas, primarily in the left hemisphere, which implement the semantics of language. It also shows the neural pointers which I propose as their connectivity. I will discuss, in sequence, (i) the relationships between working memory, short-term memory, and long-term memory, (ii) the neural functional areas identified by nTM, and then (iii) the pointers that connect these neural functional areas.

The point of localizing these functional areas, of course, is so functions can be aligned with data that associate them with specific areas in the brain, by such means as function/lesion studies, electrode mapping used in preparing for epilepsy surgery, and various forms of trans-cortical magnetic or electrical stimulation of specific locations in the brain. Besides the clinical use of such data, without it all that neuroscience could say about cognitive functions in the brain is that they are carried out holistically, all over the brain. As true as that is, if it were all we could say, neuroscience would have failed to help us refine our functional categories to help us describe human behavior in ways that "cut Nature closer to the joints".

Figure 13 shows the nTM Semantics Model of functional areas of the brain and connections among them.

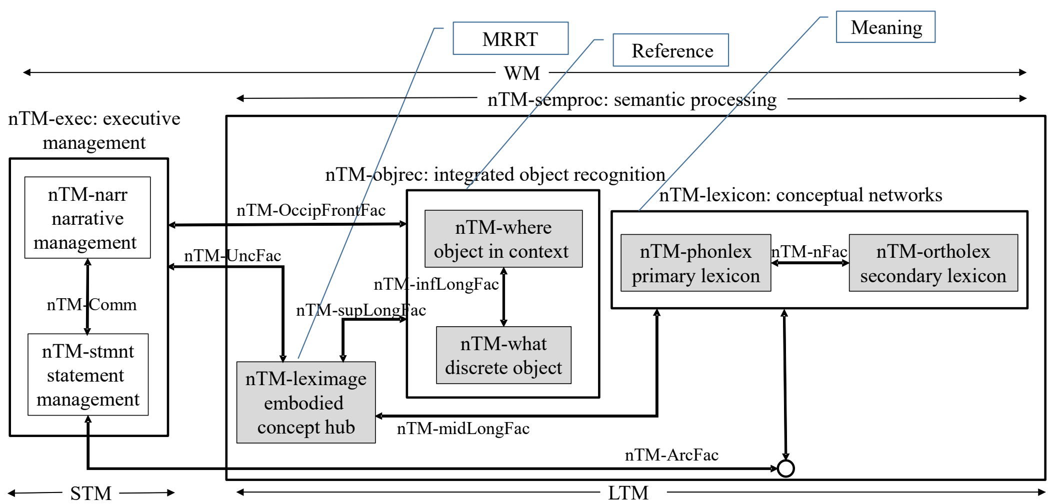

Figure 13. Neural TM: Functional Areas – Location and Connectivity.

Neural-functional areas are represented by boxes. Connections between those areas are shown as arrows. Anterior-to-posterior is left-to-right. Superior-to-inferior is top-to-bottom.

---

there is also an inability to define the kinds of identifiable things, then the lexicons have also been damaged, and the subject's linguistic behavior has declined from semantic to semiotic, from human conceptual ability to monkey sign ability. Note that these functional discriminations are similar to, but not identical to, functional discriminations that are currently used to identify classical conduction aphasias.

Working Memory, Short-Term Memory, and Long-Term Memory

Working memory (WM) is not an area of the brain. It is any above-baseline activity of a group of neurons, wherever and whenever in the brain that activity takes place. Short-term memory (STM) is activity that, in the prefrontal areas indicated as STM areas in Figure 13, does not structurally alter the neural groups involved in that activity. Long-term memory (LTM) is activity that does structurally alter active neural groups. These alterations can take the form of an increase or decrease in the number of synaptic connections on specific neurons, or changes to the spatial/temporal coordination of inputs to dendrites or cell bodies of specific neurons, or any structural alterations on either or both sides of a synapse. After such alterations to the neurons in a neural group, that neural group will respond to future inputs in a slightly different way than it would have if it had not altered its state.

WM includes, to adopt a useful distinction, both the *source* of the activities of producing and comprehending narratives, and also the *sites* of those activities. The sources of semantic neural activity are contained in nTM-exec, and are located in lateral, orbital, and medial prefrontal areas. They involve only short-term memory. The sites of those activities are contained in nTM-semproc, and are located in the more posterior neural areas that support the recognition and long-term memory of perceivable objects, and in the neural areas that support a structuralist network of concepts.

Neural-Functional Areas

For nTM, linguistic WM is the activity of producing and comprehending statements, an activity which is initiated in nTM-exec. That work is then carried out, in nTM-semproc, utilizing heightened activation elements in both the lexicons and in the object recognition and embodied concept areas. The LTM semantic processing area (nTM-semproc) is passive, in the sense that acts of linguistic comprehension or production always begin in nTM-exec. LTM is the location of the lexical and object management material which makes up the lexical items assembled and disassembled by nTM-stmnt's WM activity, in the processes of producing and comprehending statements. And both of nTM-semproc's object recognition areas (ventral/discrete-object and dorsal/object-in-context), and both of its lexicon areas (phonetic and orthographic), are updated as each statement is processed.

The narrative management area normally does not activate individual NLUs (concepts). It activates what I will call *constellations* of NLUs. A constellation is a group of NLUs that are semantically related. An NLU, to remind the reader, is what is shown in Figure 10. The semantic network of NLUs is shown in Figure 11. The basic statements produced or comprehended under guidance from the narrative management area are shown in Figure 12.

Now for the details.

nTM-exec: the Executive Management Area

nTM-exec is located in the prefrontal areas, with a (right>left) bias for narrative management, and a (left>right) bias for statement production and comprehension. It is a short-term memory function, with the memory of recently produced or comprehended individual statements lasting less than a minute usually, and the memory of recently constructed multi-level narratives fading gradually, sometimes over a period of up to several weeks. This latter timeframe is the timeframe during which we can recall an increasingly sketchy account of the *gist* of a conversation or of a chapter in a book. But the permanent record that is left by the activity of nTM-exec is in the alterations to the areas making up nTM-semproc, the long-term memory semantic processing area. Those alterations are how we learn.

As a convenient shorthand, we may think of a narrative as any series of two or more statements which are not non-sequiturs. In larger groups of statements, narratives may be nested, and may also branch and loop. nTM-exec includes a narrative management function (nTM-narr) which continually updates a multi-layered contextual representation of the information being produced by oneself or by another speaker or author, and keeps that spoken or written series of sentences aligned with an intention to pick out a set of things, and to point something out about those things, individually or collectively. As each new statement is produced or encountered, narrative management continually updates its multi-layered representation of the information being expressed.

nTM-exec also includes a statement assembly and disassembly function (nTM-stmnt). For the production of a basic statement, it assembles two concepts. The first concept picks some thing out by designating the kind of thing it is and then by including an implicit or explicit indexical to pick out the individual instance being referred to. The second concept picks out a feature which that referent instantiates.

nTM-narr: the Narrative Planning and Control Area

nTM-narr is the language management function of the executive planning and control area of the brain. It *is* executive planning and control, as applied to the job of understanding and producing statements. The evolving executive plans and controls are multi-level narrative integrations of the information content in a stream of produced or comprehended individual statements.

Isolated from nTM-narr, nTM-stmnt, the statement coding and decoding area of nTM-exec, would produce statements without any narrative coherence, and would comprehend statements in isolation from one another. Conversely, isolated from nTM-stmnt, nTM-narr would have no sequence of statements to knit together into an evolving narrative coherence.

The concepts and constellations active at any point in time exist along a continuum of activation. The most recently-used ones are the most active; less recently-used ones are less active, and at some point, ones used long-enough ago fade from short-term memory altogether, into their background resting states. For embodied concepts, there is also a most-recently-used continuum of image memory via Graded Reference Sets. For all concepts, there is a continuum of intensional memory via Graded Meaning Sets, and empirical knowledge via Graded Generalization Sets.

New constellations and concepts are introduced into WM (become active) as new referents and features are noticed. Bottom-up intrusion of lexicalized images can introduce any new concepts at all, including ones not semantically related to what has gone before. nTM-narr can decide to include those concepts in its evolving narrative, or it can ignore them. Another source of new referents and features are verb-based statements. These statements contain two or three or sometimes more thematic roles, and it is referents, with their features, that fill those roles. This top-down introduction of new material is likely to be semantically relevant, and so nTM-narr will point to new constellations based on the new concepts introduced by that input.

Executive planning and control processes impose top-down control on narrative coherence. Separate but related topics may intrude, via their associations with concepts already used. The overall linear sequence of a coherent narrative is accounted for by the ever-evolving set of most-recently-used concepts and their related concepts.

More elaborate, e.g. hierarchical, organization of narrative is also a function of the narrative management area of the brain. That coherence is managed as a nested and often branching set of plans, of objectives to pick things out and say something about them. Basic statements are the basic level of those plans.

To do all this, the nTM-narr function must have continually updated access to the series of statements produced by nTM-stmnt, and also to the constellations from which the specific concepts in those statements were selected. This continually updated access is neural *working memory*, a form of short-term memory in which the memory contents are being actively manipulated.

I place nTM-narr where neuroscience places executive planning and control functions, in anterior, lateral, and medial areas of prefrontal cortex (right>left), including dorsolateral prefrontal cortex and the anterior caudate. However, evidence indicates that nTM-narr is in fact located bilaterally, in both hemispheres. This facilitates its highly interactive exchanges with left-lateralized nTM-stmnt.

nTM-stmnt: the Statement Processing Area

nTM-stmnt is where statements are both decoded and encoded, but this work requires cooperation with nTM-narr which provides the planning and control management of an evolving series of statements which makes it possible to produce multi-statement coherent narratives, and also makes it possible to understand each statement we produce or comprehend at multiple levels of semantic coherence.

I place nTM-stmnt (not surprisingly) close to Broca's area, in left anterior and lateral prefrontal cortex, including posterior ventrolateral prefrontal cortex. But note that I regard this area as important for statement comprehension, not just statement production.

nTM-lexicon: the Lexicon Area

There are two lexicons in the brains of uni-lingual speakers. The primary one is the phonological lexicon of their language. The secondary one is the orthographic lexicon they later learn to read and write. These are independent centers of wordforms, and of the intensional relationships to other wordforms that make them concepts. These two lexicons are kept semantically coordinated by the processes described below.

nTM-phonlex: the Phonetic Lexicon

nTM-phonlex is the primary lexicon, the one developed as we first learn language. The wordforms in this lexicon consist of phonological sequences. These wordforms do not obtain their semantics from connections to an as-yet-undiscovered abstract semantic lexicon. The wordforms are connected *within* the phonological lexicon, forming a network, what TM calls the Graded Meaning Set for each lexical item. For each lexical item, there are any number of other lexical items with which it is linked. This is what makes the phonological lexicon a semantic lexicon (with the orthographic lexicon being a separately-located second semantic lexicon).

We learn these pairings as we learn our language. Some of them are so strong as to be inviolable, such as the ordered pairs [Bachelor::Unmarried], [Horse::Animal], or [Car::Vehicle]. They are inviolable in the sense that, given the ascription of the first item in the pair, if the ascription of the second item is negated or withheld, that speech act is judged to be a semantic error. A statement about cars which are not vehicles, or of bachelors who are not married, are semantic errors; those concept pairs cannot be co-ascribed to the same referent. The second item in each pair is itself the first item in another set of ordered word-pairs, and again some of those word-pairs are so entrenched that a co-ascription that negated the second one would simply be regarded as a semantic mistake.

There are also *unordered* word-pairs, ones which give a “ticket” for their co-ascription, such as [Horse:::Pasture], [Car:::Road], or [Flower:::Pot]. They may be considered an extension of Meaning Sets because while ordered pairs forbid co-ascriptional violations, unordered pairs permit, but do not require, their co-ascriptions to the same referents.

In this way, a network of lexical item constraints is built up through exposure to language. The meaning of a lexical item is the always-being-modified set of ordered word-pairs in which it is the first item, and which then requires assent to the ascription of the second item, given assent to the ascription of the first item. If we say that someone is a bachelor, we cannot deny that he is unmarried.

I place nTM-phonlex where neuroscientists place the phonetic lexicon, close to and somewhat posterior to Heschl's area, in the middle superior temporal gyrus.

nTM-ortholex: the Orthographic Lexicon

nTM-ortholex is the orthographic lexicon. For most of us, it is our secondary lexicon, the one we first learn in school. It is a lexicon of orthographic word images. Meaning Sets in this network exert the same semantic constraints as do those in the primary lexicon. In the orthographic lexicon, word-pair co-occurrences are observed through reading written text (using both the direct and indirect routes), and are manifested in the statements we write.

I place nTM-ortholex where neuroscientists place the VWFA – the visual wordform area – in the superior posterior temporal lobe, close to the parietal and occipital lobes.

nTM-objrec: the Integrated Object Recognition Area

nTM-what and nTM-where are, roughly, neural-functional areas where the recognition of (i) individual objects and their features (in a visual ventral stream), and (ii) the coordinate location of objects in space, and their relative positions and changes of position (in a visual dorsal stream), takes place. I situate the integrated awareness of objects in pointers between these two areas. nTM-what is localized in the inferior middle and anterior temporal cortex. nTM-where is located in posterior parietal cortex.

These recognitions are images which integrate cross-episodic, multi-modal, omni-perspectival, and contextualized images of objects. Single-modality images are themselves located close to the sensory-specific processing areas for those modalities. Their integration, in episodes of encounter with physical objects, is carried out by the hippocampus, but is eventually stored in hippocampal- and episode-independent long-term memory. As recorded in LTM, each episode of encounter with similar-enough lexicalized images modifies the image components of those embodied concepts, extending or constricting the scope of the perceived patterns experienced as images of the kinds of thing indicated by the associated concepts.

nTM has nothing to add to current neuroscience work on the ventral and dorsal visual processing streams, or on the integration which makes it possible to recognize discrete objects as they change over time, changes in either the features of those objects or in their absolute locations within a frame of reference, or their locations relative to one another.

nTM-leximage: the Embodied Concept Hub Area

In normal cognitive experience, object recognition is not simply the awareness of an individual physical object. It is the awareness of that object as an object of a specific kind. When neuroscientists speak about naming objects, such as responding with the word "horse" to an image of a horse, what they are talking about is *not* naming an object, but rather identifying the kind of that object. If a basic statement is being made, the subject term of the statement also includes an indexical to complete the identification of the specific object being referred to.

nTM-leximage is located in the ATL hub area of the Hub and Spoke Model. It is where words and objects are associated via pointers to, respectively, the lexicons and the integrated object area. The association is, therefore, an association of pointers. However, the associations currently identified by the Hub and Spoke Model do not include lexical pointers, because the intensional network of concepts identified in nTM is unrecognized in all current neurosemantic models, even in the Dual Coding Models which do recognize word-to-word associations as well as word-to-object associations.

It is worth emphasizing that nTM's ATL hubs are not the same kind of hubs as those posited by the Hub and Spoke Model. In the latter case, the achieved integration is apparently thought of as an integration of single-modality, and perhaps as well as of single-perspectival, images of objects. The association of wordforms with these integrated images appears to be nothing more than associating a name with the image, and since the meanings of concepts are thought to not be co-located with their wordforms, there would appear to be only a minimal amount of semantic significance to the association of a wordform with an integrated image.

But nTM finds little image consolidation work for the ATL areas to do, although in doing so, it may conflict with a substantial amount of neurophysiological evidence to the contrary. I can only hope that the data supporting those evidentiary conclusions can be re-interpreted along the lines suggested by nTM. As nTM has it, discrete object recognition, in the inferior temporal lobes, is already integrated, across sensory modalities, across episodes of encounter, and across perspectival viewpoints; and so there is no further object integration work for the ATLs to do. And in the visual dorsal stream, the ability to recognize these objects as they undergo change – change in their perceivable properties, change in their location relative to other objects which may also be undergoing change in location, and change in their location relative to fixed grid locations in a scene – is already supported by parietal areas. So I don't quite see why either static or dynamic images of objects require any neural integration work beyond that done by these two areas posterior to the ATLs.

And here is the most important point of difference of all between nTM and current neurosemantic theories. As already extensively explained – in Part 1 and above – philosophers and lexical semanticists have long recognized a distinction between word-to-object associations and word-to-word associations. Sentences which are *true by definition* are true solely because of the meaning of the lexical items which are co-ascribed in those sentences. Sentences which are *true as a matter of fact* also depend on each of their embodied concepts correctly naming the kind of object or feature that they refer to as they are ascribed to those objects and features in those statements.

Functional Area Connectivity

Turning from functional areas to their connectivity, I address connectivity within nTM-exec, connectivity within nTM-semproc, and connectivity between those two areas.

I propose that the activation of nTM-semproc by nTM-exec must be carried out by fasciculi. We may think of them as pointers between nTM-exec and the two major nTM-semproc areas – the two lexicons and the two object management areas.

> Aside: I think that there is at least a metaphorical relationship with our network of roadways (and railways, airways, telecommunications networks and, indeed, all networks in which things are moved from one place to another). In fact, I think that the relationship is more than metaphorical.
>
> Specific NLUs and NLU constellations, and/or specific integrated objects, and/or specific lexicalized images / embodied concepts, are the sources and destinations of semantic neural traffic. So LTM links must involve not only neural interstate highways, but also their exit and entrance ramps, and the feeder roads that link them, finally, to their communicating sources and destinations. But experimental neuroscience is apparently not yet advanced enough to discriminate links at anything finer-grained than the interstate highway level.
>
> A related metaphor comes from telecommunications. In this metaphor, fasciculi are nested bundles of point-to-point source/destination connections, like the separate wavelengths in a (non-switched) optical cable. In now ancient telecommunications technology, the analogy would be with bundles of twisted pairs of wires, each one contained in a (non-switched) copper-wire cable, each one a point-to-point source/destination connection. Point-to-multipoint connections also exist in telecommunications. In addition, switching, multiplexing, demultiplexing, add/drop multiplexing, dynamic routing, and stacked protocol layers, are other network concepts instantiated in telecommunications networks, and in all of the other types of networks mentioned, and applying these analogies to the brain might lead to fruitful research.

As explained in more detail below, it seems likely that (i) commisural connections relate narrative to statement management, (ii) the occipito-frontal fasciculus relates executive management to integrated object recognition, (iii) the uncinate fasciculus relates executive management to lexicalized image / embodied concept hubs, (iv) the arcuate fasciculus relates statement management to the lexicons, (v) the inferior longitudinal fasciculus relates the What and Where integrated object recognition areas, (vi) non-fascicular adjacency connections relate the two lexicons to one another, (vii) the superior longitudinal fasciculus relates the lexicalized image / embodied concept hubs to object recognition, and (viii) the middle longitudinal fasciculus relates the ATL hubs to the lexicons.

Note that executive management has both direct connections to object recognition and lexical areas and also, for embodied concepts, indirect connections through the ATL hubs.[21] The semantic tension between these direct and indirect connections influences what traces are laid down in LTM, and is another contributing factor in the extensional/intensional interplay I have called the Meaning/Reference Reciprocity Thesis (MRRT).

For linguistic cognitive functions, these pointers linking these neural areas are neural groups whose firing raises selected concept constellations and object recognitions in LTM to a heightened level of activation, and whose evolving narrative plans and statement production and comprehension processes are kept continually aware of what is going on in the LTM areas.

nTM-OccipFrontFac and nTM-arcFac: Direct Links Between Sites of Control and Sources of Activity

These two links enable direct interaction between nTM-exec and nTM-semproc long-term memory. Since one of the two areas related by each link is an exclusively STM area, these links are dynamic, carrying high-activity pulses of signals as and when the needs of narrative management and statement management arise.

LTM is the source of the concept constellations which nTM-narr blends into its multi-level narrative which (in both production and comprehension), it is constructing. And it is out of pairs of those constellations that, during statement production, nTM-stmnt selects the specific concepts, the specific words, it will use in the basic statement it is about to produce.

LTM is also, for embodied concepts, the source of recognition of (i) perceivable objects, as discrete objects and features of specific kinds, as objects in a static or dynamic context, and (ii) the LTM integration of a specific object and a specific wordform. But that indirect route, through embodied concept hubs, is carried out by two different fasciculi. A direct route to the lexicons is needed because many concepts are abstract concepts, carrying no association with images of perceivable objects. A direct route to integrated object recognition is needed because, along the network of intensional relationships branching out from abstract concepts, there may be related concepts which are less abstract, but for which the infrequency of co-ascription with an abstract concept means that ATL hub word-to-object associations have not been established for that co-ascribed pair.

[21] See (Banesh and Compton, 2018), pp. 36-37, and (Kemmerer, 2014), pp. 20-21.

During statement comprehension, access to the constellation pointers active in nTM-narr allow nTM-stmnt to be aware of the two concepts in its currently-being-comprehended statement in their constellation contexts and, even further, in the multi-layered narrative context being dynamically created by nTM-narr. And during statement production, the specific concepts chosen out of nTM-narr's constellations are modified by that new statement, strengthening the pair-bond between the two chosen co-ascribed concepts, and adjusting other entailment and inference relationships, empirical knowledge generalizations, and referential word-to-object connections as appropriate.

nTM-uncFac: the Link Between Executive Management and Embodied Concept Hubs

I propose that the uncinate fasciculus connects both narrative and statement management to lexicalized image / embodied concept hubs. Since these word-object hubs appear to have their own direct links to the recognition of objects via their integrated neural images, and to the recognition of neural wordforms with their intensional and generalization relationships within the neural lexicons, these hubs also serve as an indirect link between executive management and those two latter areas.

nTM-comm: the link Between Narrative Management and Statement Processing

This is the anterior commisure. It is a STM link within the nTM-exec area, active as long as a narrative is being constructed. In nTM-exec, both multi-level inter-statement semantic coherence, and intra-statement ascriptional and co-ascriptional coherence, are produced and comprehended. This *is* the interaction between nTM-narr and nTM-stmnt.

Statements are produced and comprehended, one by one, in the context of a fading awareness of the recent constellations of concepts and images already accessed in production and comprehension. This happens in the larger context of an always-refreshed narrative plan, and an always-refreshed set of pairs of concept constellations, provided by nTM-narr and available to nTM-stmnt.

As for statement production, nTM-stmnt uses the basic statement structure described in the *Compositionality* section above, and shown in Figure 12. Two concepts are selected, one from each of the two constellations most strongly activated by nTM-narr. If the concepts are embodied concepts, the ones selected will be ones that as uncontroversially as possible represent the perceived individuals and features that the statement is about. Whether or not the concepts are embodied, the ones chosen will be the ones which, as far as possible, introduce the minimum amount of intensional tension with previously-used words and with one another, and which bring into closer relevance, and perhaps next-sentence appearance, additional co-ascribed lexical items which best suit the intentions of the speaker for what she wants to say.

An appropriate level in a hyponymy tree must often be chosen. Most-recently-used lexical items are also given preference in the process, and this also enhances narrative continuity across sentences. One of the last steps in the work of nTM-stmnt in producing a basic statement is a syntactic step; it is to map from a hierarchical phrase-structure organization of concepts onto the linearized structure of an about-to-be-produced statement.[22]

As for statement comprehension, nTM-stmnt uses the same basic statement structure described in the *Compositionality* section above. It identifies the two concepts in the statement which will, if previously encountered, still exist in their active concept constellations, and if making their first appearances in the narrative, will have been immediately activated in the lexicon area by nTM-narr. Regardless, those two concepts will be then apprehended in both their conceptual and lexicalized image contexts. As each statement is comprehended, nTM-narr updates the multi-layered narrative it is working on. nTM-narr is working out the story, understanding it on multiple levels, and it provides nTM-stmnt with access to this evolving construction, and to statement-relevant constellations, so that the individual statement can be understood and produced in its narrative context, not just in its immediate lexical context.

> Aside: I am aware that the idea that both production and comprehension take place in the same anterior areas of the brain, is a controversial one. Much neurophysiological evidence may seem to be inconsistent with this hypothesis. But I propose that much of that evidence may come from looking at too narrow an ROI (region of interest) when gathering data on comprehension in particular, and so not including the nTM-stmnt functional area both when studying production and when studying comprehension.
>
> On the other hand, if further neurophysiological data still shows more activity in production vs. comprehension in separate areas of the brain than in overlapping areas of the brain – perhaps as precisely as function-voxel mapping using a multivariate pattern analysis paradigm could reveal – then my hypothesis would have to be modified. Perhaps I would then distinguish two kinds of comprehension, one being comprehension of concept constellations, and another being comprehension of the statement produced from or resulting from the

22 This phrase-structure tree is maximally simple, for basic statements, but can become quite complex for real-life sentences. I also note that reference to phrase-structures as the last step in statement production (and the first step in statement comprehension) is not a simple reversal of early Chomskyan theories which seem to place syntactic structures first in the statement production process, and the selection of lexical items to plug into the slots in those structures as a later step. I believe that, even in statement production, phrase-structures are accessed both early and late. They are accessed early because they provide essential semantic content over and above the content provided by constituent lexical items. That additional semantic content is, very roughly, the information about who did what to whom. And they are accessed late, of course, as the structure from which the linear sequence of words is produced as the finished statement.

combination of concepts selected from or found in each of a pair of constellations. This is an example of the way that theorizing about cognitive linguistic functions, and investigating cognitive neural mechanisms, work together to refine our functional discriminations which in turn suggest where in the brain to look next.

nTM-nFac: the Link Between Lexicons

In LTM, the primary and secondary lexicons are connected, so that semantic consistency can be maintained across them. This is shown as the one non-fascicular link in Figure 13. The two lexicons, for each of us, are kept closely synchronized, although for highly-educated people, the orthographic lexicon is usually more extensive than the phonological one.

Reading a text that someone else is speaking is something that seldom happens to adult language users. But in the early grades of primary school, it is the simultaneous reading and hearing of upper- and lower-case letters, and then of syllables, words, and sentences, that creates the phonological-orthographic link that, at the level of full wordforms, keeps the two lexicons semantically synchronized. Another way the two lexicons are kept synchronized is through the process of silently pronouncing much of what we read. This usually doesn't reach the point of silently activating articulatory mechanisms, but it sometimes does; and even when it doesn't, there is usually a level of pre-articulatory silent sounding out of what we read.

Together, these two lexicons constitute the long-term memory (LTM) of the words we learn, and of the meanings of those words. Word-pairs in Meaning Sets can come and go, in the process altering the formal and informal definitions we have for the meanings of those words. Word-pairs can move “down” the graded sets of ordered pairs, from entailments to inferences, and can even cross from inferences to generalizations, in the process becoming unordered pairs. And, of course, movement in the opposite direction is also possible. All these movements contribute to the dynamism of the Meaning Set network, and explain the phenomenon of lexical meaning change, and of the increase in empirical knowledge about the world, both over persons and groups of persons, and across periods of time.

Nothing like this appears in standard neuroscience semantic theories, with the exception of Paivio's and Vigliocco's Dual Coding Models which, as I will discuss later, have their own shortcomings as semantic models.

nTM-infLongFac: the Link Between Static and Dynamic Object Recognition

In the LTM area I call the object recognition area, and which consists of nTM-what and nTM-where as shown in Figure 13, this link represents the coordination of perception, in the visual modality, across the ventral stream constructs that eventuate in recognition of objects and features of an identifiable kind, and the dorsal stream constructs that situate individual objects in a context represented, in part, by grid cells, place cells, and time cells in parahippocampal areas of the brain. nTM has nothing to add to the research going on into this coordination of perception. Physical objects, in both their dorsal and ventral apprehensions, are the usual referents of embodied concepts, although what verbs represent can also be perceived and, in their gerundive form, referred to. nTM simply notes that linguistically-mediated reference is not simply a matter of seeing what is in front of our eyes. Sometimes intensional relationships among word-pairs may force us to reconsider. It is this influence of intensional relationships on perceptual recognition which has been entirely overlooked in current neuroscience. I believe that a LTM integration of the discrete-object and object-in-context functional areas may exist, as an evolving permanent record, in the intraparietal sulcus.

nTM-supLongFac and nTM-midLongFac: the Links between Lexicalized Images / Embodied Concepts and the Object Recognition and Lexicon Areas

In LTM, the superior longitudinal fasciculus connects embodied concept hubs with object recognition, and the middle longitudinal fasciculus connects embodied concept hubs with the lexicons. These are a Hebbian-associated pair of links whose result exists as lexicalized images / embodied concepts. Current neuroscience has a near consensus that these *hubs* relating words and objects exists in the ATLs – the anterior temporal lobes. On the nTM interpretation of these hubs, they have their work cut out for them, because they must coordinate activation states that they are in touch with via separate fasciculi. This coordination might be achieved by a pair of Hebbian-associated pointers, one to the specific object image and another to the specific wordform associated with it. It also suggests that two separate manifestations of a conduction aphasia involving the ATLs should be discoverable in the experimental evidence.

Summary of Locations and Links

Having localized nTM functional areas about where, I think, most neuroscientists would place them, I have not ended up simply where neuroscientists currently are. Here is where nTM differs from current neuroscience. The differences correspond to the four fundamental hypotheses of TM, presented at the start of this essay.

- nTM has no abstract conceptual area, where semantics exists. The meaning component of semantics is located in the phonetic and orthographic lexicons. The reference component is in the object recognition areas. Embodied concepts link meaning and reference.[23]

- nTM has no conceptual atoms. Concepts exist in, and take their meaning

[23] Of course, we can refer with all concepts, not just embodied ones. In this broader sense of reference, the first word in an ordered word-pair refers; it picks some thing out, and the second word then says something about it. And the second word refers also, to the noted feature of that thing.

from, a structuralist network of other concepts which constrain their co-ascriptive use. No concept is an island. No statement is an archipelago.

- nTM's embodied concepts are concepts in lexicons which have Hebbian-associated links to integrated images. These images have core areas, shading into penumbral areas and, ultimately, into areas which overlap with the images named by other embodied concepts. But embodied concepts are more than the names of kinds of things. Monkeys have names for kinds of things. It is only language-users that have concepts for them.

- The sensory-image-constrained referential use of embodied concepts, and the intensionally-constrained use of all concepts, control the ascriptions and co-ascriptions of our concepts to what they are about. The entailments, inferences and generalizations recorded in the lexicons control the co-ascription of our concepts in basic statements. If I am correct that any statement, no matter how complex, can be parsed into a boolean combination of basic statements which, when the syntactic "who does what to whom" information is also captured, expresses the full content of that original statement, without loss of information, then the Meaning/Reference Reciprocity Thesis explains semantics. It explains what our statements mean, what they refer to, how they are understood, and how they are produced. It explains the contributions which lexical items make to the meanings of statements, and how the use of lexical items in statements alters lexical meaning at the level of both tokens and types.

I now proceed to critique five major neuroscience semantic theories. I will explain what I think is wrong about each of them, and also how modifications to incorporate nTM can correct them.

nTM and Other Neuroscience Semantic Theories

At a high level, the relationship between nTM and current neuroscience semantic theories is this: those theories have a "semantic core", but also include either or both of the input and output streams of language, those being the mechanisms of sensory apprehension and/or motor program production.

This structure is something like an artificial intelligence core embedded in a robot with sensory and motor attachments. But nTM isn't about the robotics. I propose to leave the robotics alone, and to replace the semantic cores of these theories with nTM.

The Wernicke-Lichtheim-Geschwind Model

The Wernicke-Lichtheim-Geschwind Model assumes that there is a representation of the semantics of concepts which is distinct from both the phonological representations of their wordforms and the motor representations of the articulatory mechanisms for producing those wordforms. This is shown in Figure 14.

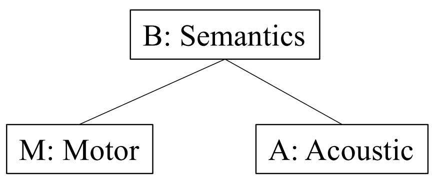

Figure 14. The Wernicke-Lichtheim-Geschwind Model.

This is famously called the "house model", with semantics being the roof of the house. This associates acoustic input with Wernicke's area and motor output with Broca's area. Other Wernicke/Broca interpretations associate semantics with Wernicke's area and syntax with Broca's area, or the meaning of words with Wernicke's area and the meaning of sentences with Broca's area.

Recent neuroscience critiques of this model focus on its imprecision. Prior to such critiques, one perspective on the model was that comprehension takes place in Wernicke's area, and production in Broca's area. A second perspective was that lexical items are managed in Wernicke's area, and sentences in Broca's area. A third perspective was that linguistic input is processed in Wernicke's area, and linguistic output produced by Broca's area. However, recent commentaries on this model agree that the linguistic functional distinctions are not as simple as this account would have it – syntax/semantics, words/sentences, comprehension/production – nor as clearly parceled out between these two distinct areas of the brain. And that is surely true. But lest we get too comfortable with the pendulum having swung so far away from these classical distinctions, let me point out the obvious corrective: there *is* a distinction between syntax and semantics, between words and sentences, and between comprehension and production.

The Wernicke-Lichtheim-Geschwind Model also emphasizes the arcuate fasciculus as the means by which Wernicke's and Broca's areas communicate. And we now know, of course, that several major tracts, including the longitudinal and uncinate fasciculi, also connect areas within the temporal lobes, and between those lobes and areas in occipital, parietal, frontal and prefrontal, and insular lobes.

I will leave these correctives to the Wernicke-Lichtheim-Geschwind Model to neuroscientists. What I will focus on is not the imprecision of that model. Rather, it is the notion that the meanings of words exist somewhere in the brain, but not where the phonological forms of those words exist.

> In the Wernicke-Lichtheim-Geschwind Model, "the center that stores the sound-based representations of words has separate links with two other components: one pathway projects to the center that contains the meanings of words, and a different pathway projects to the center for speech planning and production." (Kemmerer, 113).

But this hypothesis makes the neural realization of meaning mysterious. The brain can clearly recognize visual or auditory patterns corresponding to written or spoken words. But if the meanings of those words are not realized as relationships among those phonetic or orthographic wordforms, then how are they realized? What, neurally, are they?

The phonetic sequences representing words could certainly be stored in a phonetic network, separate from the meanings of those words. But I frankly have no idea how the *meanings* of words could be represented in another area of the brain, separate from the words themselves. If there were such a center, how would the meanings in that center be represented, if not by the words whose meanings they are? How would the neural representations of different meanings be distinguished, if not as the different meanings of different wordform phonetic sequences? Clearly the idea that I am criticizing here is not that meanings can be stored with orthographic representations instead of phonetic ones, but is rather that the meanings of words can be stored in a separate area of the brain apart from where *any* representations of the words themselves are stored.

I attribute the lack of attention to this question, on the part of neuroscientists, to their misunderstanding of meaning as reference, and to their mistaken belief that there is nothing more to semantics than reference (which they call meaning). Having accepted the Grounded Cognition meta-hypothesis, they think of the meaningfulness of a word as the association of that word with the cross-episodic, multi-modal, omni-perspectival and contextualized image of a category of objects – as the association of the wordform "tiger", for example, with the cross-episodic, multi-modal, omni-perspectival and contextualized LTM image of tigers, and they think of "meaning" as a cognate of "meaningfulness".

But as I have pointed out before, monkeys can do that. They can use their word "tiger" as a semiotic sign, and by using it on specific encounters, place specific tigers into that specific category. But they cannot use their word “tiger” as a concept because to do so is to use words in conformity with co-ascriptional word-to-word constraints. It is a simplification, but not an oversimplification, to say that neuroscientists currently operate with a *proto-semantic* theory, one which cannot distinguish human concepts from the proto-concepts used by several animal species.

Different words have different meanings. Different meanings must have different neural representations. nTM describes Hebbian-learned entailment relationships, inference relationships, and generalization relationships between a lexical item and other lexical items to which it has one or another of these relationships. The entailment and inference relationships account for the intension of the word, what Frege called its *sense*. And sense is what accounts for the fact that fluent speakers do not accept that someone who is a bachelor may also be married.

For words which correspond to things we can experience perceptually – physical objects, features and processes – nTM describes Hebbian-learned associations between a phonetic sequence for a word and a perceived, recalled or imagined perceptual image of an object. These words, which nTM calls embodied concepts, are the names of kinds of things, not of specific things themselves. A specific thing is picked out by naming a category it falls into, and then adding a lexical or gestural indexical to pick out the particular member of that category about which the predicate of the statement will then tell us something. This accounts for the extension of these words, what Frege called their *reference*. And reference is what accounts for the fact that fluent speakers do not accept that something rolling along a road on rubber tires is a rare species of orchid. We just *see* that it isn't.[24]

So for nTM, there is no “center that contains the meanings of words” that is separate from “the center that stores the sound-based representations of words”. But there is a center than contains both the sounds and meanings of words that is separate from the center that assembles multi-modal sensory patterns into the recognition of objects, their properties and relationships, and the roles they play in dynamic processes of change –

24 It's just a perceivable fact, of course, that “orchid” doesn't refer to anything that looks like that. But it is also something we know about orchids. For example, we know that orchids don't have rubber tires. That is, we know that the co-ascription of “orchid” and “having rubber tires” describes no object anywhere in the real world. Or, in order to avoid epistemological issues about knowledge, our linguistic behavior includes not accepting such a co-ascription, under any circumstances, as describing anything that we would say actually exists. And except for those with fantasy-novel imaginations, as describing no possible object whatsoever. The association of the word “orchid” with the perceptual image we have of orchids co-exists with the entailment-, inference-, and generalization-strength co-ascriptional constraints just mentioned. In the entailment case, the co-ascription evidences a misunderstanding of the meaning of either or both lexical items. In the inference case, it evidences a misunderstanding of what kinds of things in fact exist in the world. Within language, reference and meaning co-exist, and mutually constrain extensional ascription and intensional co-ascription. This is TM's Meaning/Reference Reciprocity Thesis.

change in relationships such as spatial/temporal relative locations, or change in the state of one or more objects, which is change in an object's properties.

Critique

nTM is a neural theory of semantics, and so of the “roof” of the Wernicke-Lichtheim-Geschwind Model. And there are two correctives that it applies to that model's account of semantics. First, nTM claims that the meanings of words are co-located, in the phonological lexicon, with those phonological wordforms themselves.

Secondly, nTM does not merely assert that the meanings of words are co-located with the wordforms themselves, in the phonological lexicon. It also explains what lexical meaning is, how it is different from lexical reference, and how the two interact as statements are formed from lexical items. In doing so, it is a corrective to all but one of the neurosemantic theories reviewed here.

The mistake made by those theories is based on drawing incorrect conclusions from neuroscience evidence that single- and multi-modal LTM images are active when concrete noun language is being used. These activated images are images of what concrete nouns refer to, and at the time the Grounded Cognition meta-hypothesis was being developed, it was natural to see it as a corrective to the then-dominant theory that meaning is a matter of abstract symbols being manipulated in the brain.

So the idea was that since meaning in the brain involves no *abstract* symbols, then meaning is ultimately based on *concrete* symbols, i.e. lexical items whose meaning is a yet-to-be-articulated function of amalgamated sensory perception associated with lexical item wordforms. Meaning, in other words, must be reference.

But if that is true, then there is no difference between our use of words like “tiger”, and monkeys' use of their corresponding word. Both we and they use our words to designate a specific object as a member of a category.

So something is wrong here. Something is wrong with current neuroscience reduction of semantics to reference, and something unfortunate in the neuroscience use of the word “meaning” to designate *extension*, whereas since the time of Kant (and especially since the time of Frege and Carnap), philosophers have used the word “meaning” to designate *intension*.

> *In Frege's terms, neuroscientists have ignored sense, discussed reference as meaning, and assumed that, except for the details, that is all there is to semantics.*

Commonsense intuitions about the meaningfulness of language, analytic philosophy of language work on the analytic/synthetic and a priori / a posteriori distinctions, and lexical semantics work on co-occurring pairs of lexical items, all agree that the meaningfulness of lexical items is not simply a matter of their use in referring to what we can perceive in the world around us. That is half of the picture of semantics, the half misnamed by neuroscientists as meaning, and properly named as reference. But the other half of the picture is what the term “meaning” properly applies to, and that is the set of constraints which govern the co-ascription of two lexical items to the same referent. These are the constraints which explain why a priori true statements are true regardless of what the world we experience is like. That explanation is that such statements are analytically true, where “analytically true” means “true by virtue of meaning alone”, and truth by meaning alone is how we describe our linguistic dispositions to refuse to countenance counterexamples to the co-ascriptions, no matter what.

The Lemma Model

Willem Levelt's Lemma Model (Levelt 1989, 1999a, 1999b, 2001; Levelt et al, 1999; Indefrey and Levelt, 2000, 2004) is a model of speech production, with allusions to both mental and abstract semantic structures and processes. nTM is a theory of semantics, not of the sensory input or motor output processes of language, and it considers mental and/or abstract entities as having no explanatory value in a neuroscience theory.

nTM is a neural semantic theory. To cleanly set off this narrow focus on semantics, I will, as I mentioned before, refer to all other neural groups, no matter how essential to semantic processing, as the “robotic periphery” of semantics. Of course, while the PNS can be easily conceptualized as the robotics of language, it is more difficult to relegate many sub-cortical neural groups, such as thalamic nuclei, to a secondary role. The cortico-thalamic loop certainly is involved with semantics, if only in its role of keeping recently noticed items in short-term memory. In addition, of course, this distinction exists against the reality of the holism of the entire nervous system, and especially of the cerebral hemispheres. Nonetheless, semantic core vs. robotic periphery marks an important distinction for nTM, which I propose as a substitute for the semantic cores of the other neurosemantic theories being reviewed here.

The Lemma Model suggests that the process of statement production begins by “...generating words, beginning with the formulation of communicative intentions in the realm of thoughts and feelings”. (Kemmerer, 145) In a little more detail:

> The Lemma Model “posits a complex series of computational operations that begins with an intention of what to say and proceeds … through semantic, morphosyntactic, phonological, and phonetic stages of processing. … It can be thought of, … however, as assuming an architecture that consists of two main subsystems: one for 'lexical selection', that is, identifying the most appropriate word in the mental lexicon, and another for 'form encoding', that is, preparing the word's articulatory shape.” (Kemmerer, 147)

But if “lexical selection” occurs before “form encoding”, as the Lemma Model claims it does, then where in the brain are the selected lexical items before they are form-encoded? And what does that semantic lexicon ("mental lexicon", in Levelt's terms) look like, whose components exist before they are form-encoded? This is the same mystery I described when discussing the Wernicke-Litchtheim-Geschwind Model.

It is right at the start that the mentalism of Levelt's theory is most apparent.

> The first step in word production is to map the idea one wishes to express onto a lexical concept – that is, a unit that integrates the multifarious semantic features that constitute the meaning of a particular word. This process essentially involves transforming mental states into linguistic representations. (Kemmerer, 148)

As Levelt has noted, there is indeed a "rift" in his theory. The rift he notes is between semantics and robotics. That is rift enough. But the more profound rift is in the process of "transforming mental states into linguistic representations". Even if Levelt and other neuroscientists realize that a non-mentalistic explanation of "mental states" is surely needed, it is remarkable that this mentalese persists, and has yet to be cashed in.

Like other neurosemantic theories, the Lemma Model does not recognize the word-to-word co-ascriptional constraints which constitute the meaning of those words.

> ...within the framework of the Lemma Model, lexical concepts are assumed to be represented in a non-decomposed format – that is, as unitary nodes that lack complex internal structure.... This is a controversial issue. (Kemmerer, 150)

Indeed it is. All of those concepts can be given a definition – a regimented form of which is as a boolean combination of one or more other concepts. These combinations clearly influence our linguistic behavior, e.g. in our not accepting that if "bachelor" has been applied to a person, "unmarried man" cannot be denied to also apply to him, or that if "cardiologist" applies to someone, "doctor" cannot be denied to also apply to her.[25] These constraints on co-ascriptions result in patterns of linguistic behavior that express the meanings of the involved concepts, and this structuralist network of constraints *is* the meaning of each conceptual element within that network. That network *is* the "complex inner structure" of those concepts.

The Lemma Model makes use of abstract things as well as of mental things.

[25] Examples like "horse" and "animal" are also examples of these word-to-word co-ascriptional constraints, although neuroscientists often assimilate these hyponymy relationships to the extensional/referential "concrete" words of which the hypernyms are said to be generalizations (in some unspecified sense). But [Horse::Animal] is a *word-to-word* constraint, not a *word-to-object* constraint, even though either word can be used to refer to an object being perceived or recalled. [Animal] is part of the meaning of [Horse], not part of its reference.

> ...a lemma is an abstract word node – analogous to an arbitrary number (e.g. lexical entry #2,478) that … intervenes between semantics and phonology. (Kemmerer, 151).

I maintain that there aren't any lemmas in the brain because there aren't any abstractions in the brain. Even the most abstract of concepts have their own phonetic and orthographic shapes, and except for the neural representations of those shapes, the brain has no way to record and recall those concepts. And if neuroscientists recognize this, and so talk about abstractions (and thoughts) as just placeholders in a neurosemantic theory that will eventually be cashed in, well … we are still waiting.

> Regarding … the stage that involves selecting a lexical concept that adequately represents the thought to be expressed – it is unfortunate that, as yet, very few studies have tried to discern its neural correlates. In addition, although picture naming and associative word generation … clearly recruit the core processes of conceptual preparation, they enlist this process in rather dissimilar ways. (Kemmerer, 157).

Kemmerer's first sentence here does indeed distinguish "the thought to be expressed" from "its neural correlates". My comment preceding this quote is only meant to indicate that this distinction is often elided in neuroscience discourse, and that the unfortunate consequences of doing so include a willingness to talk about "abstract objects in the brain", e.g. concepts or lemmas. But if these abstract objects aren't mental objects, what are they? And if they are mental objects, this is simply falling back on a naive and completely unarticulated version of mind-body dualism.

As to Kemmerer's second point, nTM completely agrees that picture naming and word association are different semantic processes, carried out in different areas of the brain. And nTM explains the difference. Picture naming involves lexicalized images / embodied concepts, and the ATLs are the "hub" for these embodied concepts. Word association involves word-to-word relationships, of which there are clearly semantic relationships and not just phonological ones. The two neural lexicons of nTM are the "hubs" where this takes place – in superior and posterior temporal areas (left>right), and adjoining occipital and parietal areas.

But the Lemma Model maintains that "lexical concepts are unitary nodes that … may reside in the ATLs of both hemispheres." (Kemmerer, 157) I maintain that lexical concepts are not unitary nodes, and that they do not reside in the ATLs.

However, if the only lexical concepts looked for are embodied ones, then nTM agrees that, as integrating hubs, the ATLs are where to look for evidence of them. I do add, however, that the wordforms themselves, i.e. the concepts themselves, do not reside in the ATLs. They reside only in the lexicons. What exists in the ATLs are pointers to the neural wordforms in the lexicons, and these pointers are groups of neurons that, when they fire, raise that neural wordform in the lexicons to a heightened state of activation, raising it and the wordforms it is most closely connected to out of the depths of that semantic sea to the surface of conscious awareness.[26] As these wordforms rise to the semantic surface of conscious awareness, they bring their entailment-, inference- and generalization-related NLUs along with them, and activation will spread beyond this first level of NLU-to-NLU connectivity until it fades out into the background level of activity within the lexicons. And for embodied concept NLUs, their associated images, distributed across modality-specific areas of the brain, are integrated (during formation) by the hippocampus, and eventually integrated cross-episodically, multi-modally, omni-perspectivally and contextually, in the integrated object area which the ATLs also point to.

All of this co-activation of ATLs and lexical hubs activates what I call *constellations* of concepts, of both embodied and abstract concepts, and of associated and integrated images. This is the STM raw material which nTM-narr filters down to two subset constellations, from one of which nTM-stmnt will pick out a specific concept as the kind of the referent of the statement, and another specific concept as the kind of the noticed feature of that same referent.

Statement-making will then attempt to minimize the stresses which that process will usually create within the constellations, by either temporarily or more permanently altering the always evolving image associated with an embodied concept, and/or by either temporarily or more permanently altering the intensional conceptual relationships among both embodied and abstract concepts within the constellation areas of the lexicons.

Numerous examples of this mutual accommodation in statement-making were provided in Part 1, and also here in Part 2. This mutual stress-minimizing accommodation is what the Meaning/Reference Reciprocity Thesis refers to, and this accommodation always alters the LTM "site of activity" material involved in the accommodation.

> Aside: I also note here that picture naming as a research paradigm in neuroscience is a dangerous tool because it does not distinguish between the name of a picture as a semiotic sign, and its name as a semantic concept. Monkeys can name pictures, or at least real-life presences, of such kinds of things as tigers. But they cannot name them with concepts. Neuroscientists must revisit these studies and attempt to draw this distinction, because it is, after all, the distinction between primate semiotic proto-language and human fully semantic language.

26 However, on a "copy" rather than a "pointer" theory, wordforms for embodied concepts will exist in both the lexicons and the ATLs, in which case the intensional background for a lexicalized image will become available in awareness because the similarity of the copies enables the ATL hubs to access that intensional background.

At (Kemmerer, 166), Figure 6.12 shows a schematic of both the Lemma Model and a competing model, the Independent Network Model. What both models have in common is that in the process of producing statements, we start with “semantic representations” and end up with “phonological representations” and “orthographic representations”. The schematics also show the two theories agreeing that, after a semantic representation has somehow appeared, but before it is given phonological or orthographic form, “syntactic representations” are formed.

But we have no account of what these semantic and syntactic representations are. nTM has no place for representations of abstract and/or mental objects other than their representations as how words sound or how they look on the page. And as I have tried to show, nTM has no need for abstract or mental representations in its neuroscience explanations. They are data to be accounted for, but not pieces of the neural machinery than accounts for them.

> Aside: once again, we need to look more closely at the words we use to describe the neural basis of language. In one sense, “semantic representations” could mean “neural representations, independent of the wordforms they give meaning to, that manifest themselves in meaningful linguistic output”. For neuroscientists who are comfortable with this sense of “semantic representation”, such as Levelt and proponents of most other neurosemantic theories being reviewed here, I have two questions. First, since they claim that “semantic representations” do not exist in the brain as phonetic or orthographic representations, but simply wear those representations as the clothes they put on when they prepare to go out in public, where are the unclothed semantic representations in the brain, and what are they? Second, since they accept that concepts must exist in the brain as nothing other than neural objects in the brain (and how could they deny that?), why do they so assiduously avoid accepting that the neural representations of wordform sounds and shapes are those neural objects?
>
> On the other hand, we might identify a second sense of “semantic representation”, which would go something like this. Clearly, concepts aren't just neurons. They are units of meaning, not biological cells. We already have a pretty good idea of how phonetic and orthographic wordforms are stored in the brain. But that isn't enough to explain language. As John Searle argued, we aren't robots that input and output sequences of sounds and shapes (Searle, 1980; Hauser, 2018). We are human beings communicating with one another. We use our brains to share our *ideas*.
>
> This second sense of “semantic representation” is one that, I think, would be very congenial to philosophers such as David Chalmers, who introduced “the hard problem of consciousness” to neuroscientists. For Chalmers, consciousness is a phenomenological thing, something like our awareness of things. And for

Chalmers-inclined neuroscientists of language, concepts are also phenomenological things, something like ideas that we combine to say something we want to express. And how can an account of what goes on among neurons in the brain even begin to explain that phenomenological fact of having ideas, and of using concepts to express those ideas?

For Chalmers, there is something more than neural processing going on when we are aware of things. That something more is consciousness. And on this second sense of "semantic representation", there is something more than neural processes going on when we produce a statement. That something more is the thought we have in mind, and then express by means of that statement.

For neuroscientists more attracted to this second sense of "semantic representation" than to the first sense, there is a reason that they haven't yet found a neural group that accounts for semantic representations. That is because, on their own principles, they *can't* find any such thing. Anything any of them might propose as a candidate would face the same objection already raised to any neural process hypothesis about consciousness, i.e. that concepts are units of meaning, something more than units of electrochemical connectivity.

Probably most neuroscientists of language don't even notice this distinction between these two senses of "semantic representations". And for those that do, I suspect, most of them pay little attention to it. They are getting on with their research and their theorizing, and such distinctions as this one can be left to … well, to analytic philosophers of language.

Neuroscientists are not analytic philosophers of language, trained to make and to be aware of conceptual distinctions which are not ready-to-hand. But perhaps I have shown how steep a price has been paid for not being aware of this particular distinction. It has led some neuroscientists on a wild-goose chase for thoughts, intentions and meanings which are somehow felt to be more than neural, but which are at the same time thought to be located somewhere in the brain, apart from the neural representations of the words whose meanings they are.

This is the Hunting of the Snark. nTM says the snark was right under our noses all the time.

Critique

The basic mistake of the Lemma Model is the assumption that, in statement production, "something mental" happens first. Words are then recruited to express that something mental. This is no innocuous "placeholder" use of reference to mental things like thoughts or intentions – although Kemmerer does attempt to present it as nothing more than that. But the evidence for its being more than placeholder language is right in the “happens first” assumption of the Lemma Model.

The semantic units of the thoughts we wish to express are statements, and the semantic units of statements are lexical items, and lexical items are represented in the brain as the neural correlates of phonetic wordforms. Why would Levelt and other neuroscientists want to place the neural activation of those wordforms downstream from the process of having a thought to express? Why must thoughts come first? *How* can they come first?

I think the following notion is involved in that assumption. Thoughts must cause our linguistic behavior because if they didn't, we'd just be automatons producing sounds in response to neural inputs to neural states. Causes must exist before they can produce their effects, so thoughts must exist before we can create strings of words to express them.

But for a neuroscientist, this is a self-defeating line of thought. Talk about thoughts coming first *must* be placeholder talk, Kemmerer suggests; for otherwise neuroscience explanations will invoke non-physical entities as causes of physical changes. No neurophysiological account can posit thoughts and ideas as causal elements in neural processes.

So what are thoughts placeholders for? If they are not placeholders for the neural representations of phonetic wordforms, they must be placeholders for the neural representations of something else. What else? Perhaps a distributed pattern of neural activity sequentially or concurrently activated across several distinct areas of the brain, and firing in different frequency bands over the time it takes to express a thought. But whatever the suggestion may be, the reason that thoughts are assumed to precede recruitment of the words that express them also applies to these neural representations, whatever they turn out to be. Thoughts must precede these neural representations too, because thoughts must cause them. Otherwise, once again, we become mere automatons.

In our folk theory of mind, thoughts certainly do cause us to say things. But this is not a sense of causality that has any place in any physical science. It is more like the notion of having a *reason* for saying things.

In nTM, the meanings of words are not stored separately from the neural form of those words. A concept is not retrieved before its wordform is. Rather, the neural anchor of a concept is its phonetic or orthographic physical wordform, as encoded in neurons and, existing within a network of mutually-constraining wordforms, is the concept itself. Neither abstractions nor thoughts are components of any neurophysical explanation.

Thoughts are indeed one of those things that a cognitive neuroscience must account for, but they are not one of those things that can be used in the account. As for abstractions, they are no better than the abstract theories which give rise to them. Chomsky's path through linguistics is littered with the detritus of abandoned theories, each with as many abstractions as the fertile minds of his students could dream up. I think lemmas are such an abstraction.

The Dual Stream Model

The two streams of Hickok and Poeppel's Dual Stream Model (Hickok and Poeppel, 2007; Poeppel et al, 2008) – the ventral and dorsal streams – do not correspond to TM's distinction between meaning and reference. The ventral stream, called the "What pathway", is said to implement the meanings of words while the dorsal stream, called the "Where pathway" (or, more recently, the "How pathway") supports the production of words. Their What pathway is what I'm interested in.

As a variation of the Wernicke-Lichtheim-Geschwind "house" Model, the Dual Stream Model suggests that the phonological forms of words are stored in one area of the brain, and that the meaning of words are stored in another area (or, perhaps, "everywhere"). I have already raised objections to these hypotheses.

The Dual Stream Model also incorporates a robust mentalism, just as does the Lemma Model. In their 2009 paper, the authors state that "the computational transformation between thought and acoustic waveform (and vise versa) is a complicated, multistage process." (p.122) My critique of the mentalism in the Lemma Model applies here, also. Either thoughts are mental things, in which case they play no causal role in neurophysiological processes – since only a neurophysiological process can cause another neurophysiological process to occur – or they are neurophysiological things, states and processes occurring in the brain. But in that case, where in the brain are they? And what kind of neural processes are they?

I propose that, as the title of Section 1 of their paper indicates, thoughts and thinking with language is "what we are trying to understand". Fair enough. But thoughts and thinking are not elements in any neurophysiological explanation, elements in chains of causes and effects leading to the production (or comprehension) of language.

In reviewing Hickok and Poeppel's model, Kemmerer notes that "The semantic structure of words, however, are thought to be widely distributed across a variety of cortical regions." (Kemmerer, 130). Of course, if the only words are embodied concepts, then they would indeed be widely distributed, since the single-modality contributions to their multi-modal integrations are widely distributed. If something more than that is intended by "widely distributed", then this is simply hand-waving in the direction of neural holism. Of course the brain is a holistic network of neurons. Of course everything is related to everything else. But in this sense, everything that neuroscience studies is something that occurs everywhere in the brain. Holism is a truism. It is also a singularly uninformative one as far as neuroscience explanations are concerned.

Kemmerer also discusses how Levelt's concept of a lemma might be incorporated into

Hickok and Poeppel's Dual Stream Model. He distinguishes between a one-stage and a two-stage version of lemmas. On the one-stage view, "...the meaning of a word (e.g. the concept of a cat) projects to a composite word-level phonological representation" (p.131), which in turns projects to the levels of the phonological network that spell out the phonemes and syllables in the word.

In this account, there seems to be an implicit notion of an initiating impulse "projecting" a concept from its abstract or mental existence out to its representation in sound. But nTM maintains that there is no non-phonetic concept anywhere in the brain, as either a mental or an abstract object. For nTM, there are just phonetic sequences in the lexicon, some of which constitute wordforms and have Graded Meaning Set relationships to one another. So subtract the notion of abstract entities in the brain from Hickok and Poeppel, and lemmas have no role left to play in the Dual Stream Model.

In discussing the Dual Stream Model, Kemmerer notes that:

> … there are now good reasons to suppose that the ATL contributes not only to the transient formation of the message-level representations of multi-word expressions, but also to the long-term storage of single word meanings. (Kemmerer, 134).

I fully agree. This is precisely what nTM explains. The ATLs are active in working memory. Feedback loops from STM sources of control to LTM sites of activity have already been noted, a process in which, under the guidance of executive plans and controls, WM heightens the activation of specific concept constellations in LTM and, as new statements are comprehended or produced, and as narratives evolve to make contextual sense of them, those same constellations, in their LTM lexicons, are themselves constantly evolving. In this way, nTM explains how the "transient formation of the message-level representations of multi-word expressions" (in nTM-stmnt), interacts with "the long-term storage of single word meanings" (in the nTM lexicons). The ATL contribution to this working memory process is in its management of well-learned word-object associations, resulting in the LTM of lexicalized images / embodied concepts (aka concrete words), in which the intensional constraints on concepts are associated with the images of the things we talk about.

Critique

nTM is a replacement for the semantic component of the Dual Stream Model. It has nothing to say about the dorsal stream, but it splits the ventral stream into two components -- meaning and reference, implemented as localized intensional relationships among NLUs, and distributed extensional relationships among integrated images.

nTM differs from the Dual Stream Model on two important points. One is that it abjures the notion of abstract concepts. The second is that it finds that the Dual Stream Model –

like several other neuroscience models of semantics – recognizes only reference as what semantics is about. Again, this may be a consequence of the emphasis of the Grounded Cognition Model on the association of semantics with concrete images, in contrast to the Amodal Symbolic Model which regards associations between words and images as ancillary to semantics.

If the semantic core of the Dual Stream Model is replaced by nTM, both these mistakes are avoided. In nTM, meaning exists in the lexicons and reference exists in the object recognition areas, and the lexical/object integration in the ATLs brings them together as we speak about things in the world around us. The processes of making and comprehending individual statements and their contextualizing narratives alter integrated images, NLUs, and their word-object associations, thus accounting for each individual's contribution to the evolution of her language.

The Hub and Spoke Model

> According to a theory of semantic knowledge called the Hub and Spoke Model, the ATLs are integrative regions that have bidirectional connections with each of the anatomically distributed modality-specific systems, as well as with the systems that subserve the phonological and orthographic representations of words (Figure 10.9; e.g. Rogers et al, 2004; Patterson et al, 2007; Lambon Ralph & Patterson, 2008; McClelland et al, 2009; see also Simmons & Barsalou, 2003). (Kemmerer, 286)

The Hub and Spoke Model is a model of lexicalized images, i.e. embodied concepts, often referred to as concrete nouns by neuroscientists. As such, it is an incomplete semantic model, one addressing only peripherally those lexical items which have no associated perceivable images. And like all the other neurosemantic models reviewed in this essay (including the Dual Coding Models), it is also incomplete because it recognizes only the extension of words, not their intension, only their reference, not their meaning.

> The Hub and Spoke Model is a more detailed account of "… the theory (that) maintains that conceptual processing amounts to recapitulating modality-specific states, albeit in a manner that draws mainly on high-level rather than low-level components of the perceptual and motor systems." (Kemmerer, 275)

So conceptual processing is "recapitulating modality-specific states, albeit in a manner that draws mainly on high-level rather than low-level components of the perceptual and motor systems." But that phrase is vague enough that we shouldn't try to figure out what it means. Instead, we should look elsewhere for clarification. Nonetheless, it is already clear that, for this neurosemantic model, "conceptual processing" involves nothing like what philosophers and lexical semanticists call meaning. It involves reference to objects, features and processes we become aware of through our senses.

Antonio Damasio appears to share this reduction of semantics to reference, and Kemmerer quotes him in the chapter describing the Hub and Spoke Model.

> In short, a wide array of representations will be generated that together define the meaning of the entity, momentarily … The mechanism that permits co-activation of representations depends on devices I have called convergence zones, which are ensembles of neurons that 'know about' the simultaneous occurrence of patterns of activity during the perceived or recalled experience of entities and events. (Damasio, 1989; in Kemmerer, 276).

What might Damasio mean here? First of all, to say that entities have meaning$_{dm}$ shows that meaning$_{dm}$ is a variant use of the word "meaning" which removes it entirely from the sphere of philosophy of language, linguistics, lexicography, and our ordinary folk theory understanding of what meaning is.[27] Entities that have "a wide array of representations", i.e. the things we can perceive and refer to, don't have meaning. Words do and, derivatively, sentences do.

To make the best we can of what Damasio said, I suggest that meaning$_{dm}$ be redirected to words, specifically to words that refer to things we become aware of through our senses. These are the words usually called concrete nouns. This helps Damasio along by redirecting his notion of meaning so it, like the accepted sense of the concept, is about words, not about what words are about.

But the word "dog" *is* about dogs. Its use, on any occasion, is semantically correct if and only if it is being used to refer to a specific dog, or to dogs in general.

This won't help Damasio much, though. It is essentially the naive picture theory of meaning which was briefly popular among the logical positivists of nearly a century ago. It does not apply to abstract words, ones which do not refer to perceivable things. It does not explain the relationship between definiendum and definiens which, expressed as a statement (e.g. for all x, x is a bachelor if and only if x is an unmarried adult male human being) is an analytically true statement, true no matter what sense experience presents to us. It does not explain why we sometimes must conclude that, when describing something we seem to just see, we can be judged mistaken by others, including other fluent speakers having the same sensory inputs.

With respect to Damasio's convergence zones, I believe that cross-episodic, multi-modal, omni-perspectival and contextualized long-term memories of perceivable things and events do exist, but I don't think that they are dynamic processes, happening "momentarily". I follow the lead of most neuroscientists who seem to believe that these

27 I will use "meaning$_{dm}$" when I'm talking about meaning in Damasio's sense.

convergences are structurally-encoded long-term memories which exist in the integrated object area, and are pointed to by the ATLs.

For example:

> According to the Hub and Spoke Model, object concepts of the kind typically encoded by concrete nouns are subserved not only by modality-specific brain systems for perception and action (the spokes), but also by an integrative amodal system that resides in the ATLs bilaterally (the hub). (Kemmerer, 295)

Aside from the suspicious use of "amodal", instead of "multi-modal", this point isn't relevant to an analysis of the Hub and Spoke Model from the perspective of nTM. What is relevant is the idea that the way we get to meaningful concrete words is to (momentarily or not) pull together their single-modality distributed memories into a single integrated recollection of what the word is about.

This idea, of course, is consistent with the Grounded Cognition Model. But when neuroscientists threw out such Amodal Symbolic Models as Fodor's Language of Thought, they threw out the baby with the bathwater. They threw out meaning, as the higher-level contexts of discussion shown in Figure 1 explain meaning to be. This left them with only a simplistic version of reference in which we will correctly use words to refer to things as long as our senses are unimpaired, we are fluent speakers of our language, and our intention is to say what we believe to be true. They then confused terminology by calling what they were left with "meaning". But meaning is precisely the semantic complement of what they were left with, the semantic complement of reference.

The "conceptual hemianopsia" I mentioned earlier is clearly evident in the Hub and Spoke Model, if Kemmerer is correct that the Hub and Spoke Model "maintains that it is precisely this evocation of perceptual and motor representation that constitutes the bedrock of comprehension." (Kemmerer, 286)

Kemmerer discusses how the Hub and Spoke Model interprets data about semantic dementia. He says that, in semantic dementia:

> From the perspective of the Hub and Spoke Model, the amodal hub is disrupted first, and then the visual spoke begins to malfunction as the atrophy spreads into more posterior parts of the inferior and middle temporal gyri. (Kemmerer, 289)

From the perspective of nTM, this account fails to distinguish several kinds of dissociation between words and images. In the following, I use the word "horse", and pictures of horses, to illustrate. In all cases, unless otherwise noted, damage in one area is assumed to take place in the context of intact cognitive function in all the other areas.

nTM also proposes that both narrative and statement management have direct connections to the lexical and object recognition areas, connections which do not rely on the ATLs and their hubs. And, to repeat a point I made earlier, I do not believe that the ATL hubs carry out the cross-episodic, multi-modal, omni-perspectival and contextual image integrations that I have already assigned to the integrated object recognition area.

Rather, the ATL hubs, as far as nTM is concerned, are where *well-learned* associations of words and objects are stored in LTM. Novel connections, using the direct routes shown in Figure 13, could create an initial LTM trace of their association in the form of a new ATL hub. Thereafter, relying on the ATL hubs will be more efficient for narrative and statement management because the connection between the two pointers will have already been created and, usually, will have been subsequently reinforced. Nonetheless, the existence of both a direct and an indirect connection of words with objects makes the localization of conduction aphasias, in particular, more difficult.

Finally, I will not consider the issue of relating data from multiple disrupted sites of semantic activity. The combinatorial possibilities would make that too long a discussion in this current context.

I begin with lexical damage. This is lesion- or degeneration-based damage to the lexicons. One kind of lexical damage I will call *semantic damage*. It will manifest like this. The word “horse” will remain in the lexicon, but some or all of its intensional and/or generalization connections to other words will be lost, and so the subject's ability to say what a horse is, or to relate factual knowledge about horses, will be impaired. But the subject will still be able to respond with the word “horse” to a picture of a horse. So this manifestation of semantic dementia can occur even when the ATL area is undamaged.

A second kind of lexical damage I will call *semiotic damage*. In this case, even the word itself is lost from the lexicon. It will manifest like this. The subject will be unable to associate the word “horse” with pictures of horses, or to define the word, or to evidence any factual knowledge about horses. But the subject will be able to recognize different pictures of horses as pictures of “the same thing”. She may also be able to utilize another word in the same semantic density neighborhood as the word “horse”, perhaps a hypermyn or even a hyponym. So this manifestation of semantic dementia can also occur even when the ATL area is undamaged.

The second major category of damage, relative to the Hub and Spoke Model, is object recognition damage. One kind of object recognition damage is what I will call *dynamic object recognition damage*. This is damage in the nTM-where area. It will manifest like this. The subject will be able to recognize and name pictures of horses, because these are images of static objects. But the subject will not be able to recognize or name movies of horses, because those are images of objects changing absolute locations within a fixed scene and also locations with respect to other objects in motion. So this manifestation of semantic dementia can also occur even when the ATL area is undamaged.

A second kind of object recognition damage I will call *static object recognition damage*. This is damage in the nTM-what area. In this case, it might be impossible to use the word "horse" to designate any object, whether at rest or in motion. But another possibility is that the subject would be able to identify horses in motion, but not at rest. So this manifestation of semantic dementia can also occur even when the ATL area is undamaged.

A third major category of damage is damage confined to the ATL hubs. This damage dissociates a hub's pointers, separating a pointer to an object from a pointer to the word which says what kind of object it is. It will manifest like this. The subject will be able to tell that pictures of horses are all pictures of "the same kind of thing", but will be unable to use the word "horse" to designate what kind of thing that is. In addition, the subject will be able to say what a horse is, and also to display a range of empirical knowledge about horses.

I don't wish, at this point, to describe functional / neural damage correlations any further. The overarching point, here and in my discussion of all other current neurosemantic models, is that the Grounded Cognition Model has led neuroscience astray, and created the conceptual hemianopsia I have already mentioned. The meaning of words is distinct from the reference of words, and it does not exist as some kind of mental structure and process, either in a yet-unspecified area of the brain, or somehow non-localized, being a function of the brain in its entirety. nTM is a neurosemantic theory which both localizes and explains the meanings of words, and shows how that set of semantic constraints interacts with the referential constraints of our embodied concepts / lexicalized images.

However, I suspect that little or no attention has been paid to distinguishing these different manifestations of semantic dementia, since the distinctions depend on this new neurosemantic theory. But I predict that if experimental neuroscientists will distinguish these functional impairments, and will look for the degenerative (or lesion) correlates I have described, they will find them. They will then have distinguished several hitherto unrecognized forms of semantic dementia. Perhaps the most important distinction they will discover is that between the semiotic proto-language abilities of monkeys (and of damaged humans), and the fully semantic abilities of cognitively intact human beings.

Kemmerer goes on to describe a semantic dementia study of patient EK. (Bright et al, 2008). (Kemmerer, 289ff.) In the study, tests were given to EK for three years, one year apart. There were four category of tests: (i) object naming; (ii) word-to-picture matching; (iii) category fluency; and (iv) property verification.

As interpreted by nTM, *object naming* tests the link from a lexicalized image to the wordform in the lexicon which lexicalizes that image. *Word-to-picture matching* tests the link in the opposite direction, from a word in the lexicon to the image it lexicalizes. *Category fluency* tests hyponymy relationships in the lexicons, and *property verification* tests the generalization relationships in the lexicons. For nTM, all four categories will activate both object integration and lexicon areas, but the last two categories will more strongly activate lexicon areas than object integration areas.

Even at the initial first-year test, in which damage was confined to the ATLs only, object naming was only 20% of normal. Word-to-picture mapping was 90%, but that is still below normal. The nTM interpretation of the object naming deficit is that, within the ATLs, the paired link from the object image to its associated wordform image was damaged. Word-to-picture mapping was relatively well-preserved because the pair of direct links, mediated by narrative and statement management, between the NLU in the lexicon and the image in the object integration area, was well-preserved.[28] Nonetheless, the difference in impairment between object-to-word and word-to-object association is puzzling, and I currently have no explanation for why executive management would support the latter better than the former.

Even at this early ATL-only stage, performance was "quite impaired" on category fluency (only 7 items) and property verification (72%). This indicates damage in the lexicons, where entailments, inferences and generalizations reside. Since this lexicon damage was apparently not noticed in the experiment, I suggest it is because it was not looked for, or because lexical damage, at that point in time, had not affected enough NLUs to be noticeable with current technology. The semantic lexicons of nTM are invisible to all current neurosemantic theories (other than the Dual Coding Models).

At second-year testing, damage extended from the ATLs to the middle temporal gyrus, and performance was worse on all tasks. The nTM-phonlex area was damaged, with consequent degradation to all conceptual relationships, including those involving embodied concepts.

At third-year testing, atrophy was even more extended. Category fluency declined, but word-to-picture matching remained stable. This latter point is interesting. It indicates that, via the direct pointers from executive management, relationships between the lexicons and the object integration area were relatively well preserved compared to entirely intra-lexicon word-to-word, concept-to-concept links.[29]

[28] Note that nTM does not posit a direct fascicular link between the lexicons and the object recognition area. The existence of one would certainly alter the interpretation of lesion/function data, and there is no reason why nTM could not accommodate this additional link should it be found to exist.

Note also that this is the first mention of neural semantic pointers being bidirectional. At the level of individual axons, of course, the connection is unidirectional. But, perhaps not surprisingly, at the level of fascicular tracts, it is usually bidirectional connections that are laid down and maintained.

[29] I note that monkey signs work in both directions also. So far, I have emphasized the production of those category words in the presence of tigers. But since those calls are comprehended as warnings by nearby

> Aside: neuroscience focus on reference does include hyponymy (category fluency) relationships, which neuroscientists apparently think are adequately accounted for as generalizations from more specific names of kinds of objects. But since the processes by which hyponymy relationships are created and accessed have not, to my knowledge, been described in any neurosemantic theory, the assimilation of this specific kind of lexicon entailment relationship to a semantics-as-reference adjunct, is not plausible; and I think it is indeed mistaken. Hyponymy relationships are word-to-word relationships, not word-to-image relationships.

Kemmerer goes on to discuss an experiment involving synonymy judgments.

> (See Binney, 2010, on synonymy judgments.) “Left>right dominance was observed “...perhaps because the semantic task in this study drew more heavily on lexical relations that are likely to be left-hemisphere dominant”. (Kemmerer, 294)

Well, yes. I've been saying this all along. The lexicons are in the left temporal lobes.

Patients with semantic dementia, Kemmerer states, are “gravely impaired on the synonym judgment task”. (Kemmerer, 294) nTM's interpretation of this data is as follows. Synonym judgments are pairs of entailments. Each entailment is an if-statement, e.g. “for all x, if x is a bachelor, then x is an unmarried man”. Since “bachelor” and “unmarried man” are synonyms, we have the symmetric entailment “for all x, if x is an unmarried man, then x is a bachelor”. The entailments involved in synonym judgments exist entirely within the lexicons, and nTM contains a detailed account of the lexicons, which is far more informative than a reference to “lexical relations”. I have located the secondary lexicon in the VWFA, and the primary lexicon with the phonological network, and these are more posterior than anterior areas, affected only as dementia extends itself posteriorly.

I now review three questions at the end of Kemmerer's review of the Hub and Spoke Model. First:

> Are there hemispheric asymmetries in the functions of the ATLs, perhaps with the left-side structure contributing more to linguistically than non-linguistically encoded concepts and the right-sided one having the opposite profile?

Yes, there are asymmetries. But the linguistic asymmetry is this. Narrative cohesion is managed by nTM-narr, and it resides bilaterally, but with right-side dominance. Statement production and comprehension is managed by nTM-stmnt, and for it, the left

monkeys, for them the category word must link to the same generalized experience of tigers. Both semiotic sign production and comprehension exist in primate proto-language.

hemisphere is strongly dominant.

Second:

> Do the two supposedly key regions within the ATLs – namely the anterior fusiform region and the inferolateral region – carry out different kinds of semantic operations?

Yes, they do. The anterior fusiform region, via the uncinate fasciculus, contributes to the work of nTM-narr in managing the production and comprehension of narrative coherence. But I have proposed that nTM-narr itself is in STM prefrontal areas, not in the ATLs. The inferolateral region contributes to the work of nTM-stmnt in producing and comprehending statements. But I have also proposed that nTM-stmnt, like nTM-narr, is in STM prefrontal areas, not in the ATLs. And, to give syntax a mention in this essay on semantics, it is also the region where the hierarchical phrase-structures which are the form that all our statements initially take, are translated into and out of the sequential, linearized structures of the sentences we produce and comprehend.

> Aside: it seems that most narratives, existing in STM, fade away after they have done their work. We don't remember the course of most of our conversations for very long. However, the major folk theories of mind, the major scientific theories in biology, chemistry and physics, and the great religious and mythical narratives, do seem to survive in LTM, shared among members of a linguistic community. I think that these multi-level and branching narratives survive as adjustments to mega-constellations of concepts and associated images. But of course, all this is mere speculation.

Third:

> ... how do the hub-like conceptual functions of the ATLs relate to the apparently similar hub-like conceptual functions of the left angular gyrus? (Kemmerer, 295)

The hubs in the ATLs relate embodied concepts to the images they lexicalize. They do so by means of links to and from the lexicons and the image recognition area. The hubs in the left angular gyrus – as well as in the superior mid/posterior left temporal lobes – use phonetic and orthographic wordforms as the sensorily-available anchors for the structuralist network which turns those wordforms into concepts.

Critique

The Hub and Spoke Model is another model of semantics without meaning. Names of (kinds of) objects have no more semantic content, given this model, than do a monkey's words for tigers or snakes. What is lacking is the role of the lexicon, where words are associated, in co-ascriptional constraints, with other words, knitting together a structuralist network of constraints that constitutes the meaning of all the lexical items – not just the ones that lexicalize images – that make up human language.

Nonetheless, there is increasing evidence and theoretical consensus that, in the ATLs, the integrated images of perceivable objects, and the words that name their kinds, are associated. For nTM, this is where lexicalized images / embodied concepts reside, as pointers to image recognition areas and to lexicons, and I have no objection to calling those image/concept associations hubs. However, as discussed earlier, these are not hubs in the sense that the Hub and Spoke Model understands hubs to be.

The Dual Coding Model

According to the Dual Coding Model:

> A key aspect of the verbal system is that, rather than representing word forms in complete isolation from each other, it captures complex networks of frequency-based associations among them, (including) their statistical tendency to co-occur in the same discourse contexts....far from being semantically irrelevant, these reciprocal links are assumed to actually constitute, to a large extent, the meanings of the words. (K, 337)

Vigliocco has developed a variant of Paivio's original Dual Coding Model. In her variation, there are two main systems:

> … first, an 'experiential' system that stores long-term modality-specific representations; and second, a 'distributional' system that registers the statistical co-occurrence patterns of words across discourses. … (This approach) brings together under a single rubric a substantial amount of psycholinguistic and computational data... (K, 338)

One way of approaching the Dual Coding Models is to ask two questions. First, do these co-occurrence patterns explain what words mean? An affirmative answer to this question has been called The Strong Distributional Hypothesis. But if they do, how do they do it?

Second, do the meanings of words explain these co-occurrence patterns? An affirmative answer to this question has been called The Weak Distributional Hypothesis. But if they do, how do they do it? (Lenci, 2008; pp. 14-17. Lenci, 2014; pp. 32-33)

TM, as developed in Part 1 and here in Part 2, is a fully-developed semantic and neurosemantic theory which answers both questions in the affirmative, and which, for each question, explains how it's done. Asking which one explains the other has elicited spirited defenses of both the strong and weak versions of the Distibutional Hypothesis (see Lenci, 2008). I think that's because it is an easy conceptual slide from “x explains y”, to “x accounts for y”, to “x causes y”, and from the last of these it would seem to follow that x can exist without y but must exist before y can exist.

From one of these two perspectives, statistical co-occurrences of word-pairs cannot function as a foundational given in a semantic theory. For example, why do “horse” and “pasture” co-occur? Why do “horse” and “animal” co-occur? Surely they do so for different reasons. This is something that needs to be explained. TM's explanation begins by emphasizing that co-occurrence is too wide a data-gathering net to cast. It is co-ascription we must focus on.

But it is clearly true that these statistical co-occurrence patterns do influence meaning. We revise dictionaries because patterns of the use of lexical items change over time. The discourse produced by each of us exhibits co-ascriptions of pairs of words to the same referents. TM maintains that these co-ascriptions may start off life as generalizations, but that as they continue to be used and also accepted in an ever-wider range of scenarios, by an increasing number of relevantly fluent speakers, they may take on the status of inferences. Eventually, after their inferential use has become widely-enough established, for a long-enough period of years, lexicographers may get their hands on this data and revise dictionary definitions to conform to these new patterns of use. That is the official societal seal of approval for what a word means. The co-ascription of that ordered pair of words now expresses part of what the first word means. (See Part 1, pp. 26-29).

So the conundrum of statistical co-occurrence patterns as both needing to be explained by a theory of meaning, and also as being themselves foundational to a theory of lexical meaning, is explained. Co-occurrence is too broad a net; co-ascription is what is needed. Sometimes what binds two concepts together in a co-ascription pair is reference; we can just see that the two referents are right out there, together. At other times, what binds two concepts together is the learned co-ascriptional pattern that given that something is [A], it can't be denied that it is also [B]. It's nothing in the world that gives force to this “can't” except our societal refusal to countenance such linguistic behavior. “If x is a bachelor, then x is not married” is neither an empirical deduction from an overwhelming amount of evidence, nor a report back from some realm of pure ideas.

To illustrate the importance of adding a missing meaning/reference distinction to neurophysiological studies of word comprehension and production, I will focus on how the Dual Coding and Context Availability Models account for the different ways that abstract and concrete words are processed in the brain.

In nTM's terminology, the difference between abstract and concrete words is the difference between abstract and embodied concepts. The principal difference between embodied and abstract words is that embodied words have direct physical referents, whereas abstract words have only indirect ones, via their graded entailment-, inference-, and generalization-relationships to a network of other concepts.

For abstract words – such as most of those being produced by me right now – my physical surroundings are not very important. Hence, concept-object referential associations are not particularly relevant to my choice of the abstract words I am using. So, lacking extensional set membership associations (objects to categories), abstract word choice is governed primarily by intensional set inclusion associations (categories to categories).[30] Individuals – members of sets – which are referents in statements I will call *material referents*. Concepts – sets including and included in other sets – which are referents in statements I will call *formal referents*, referents not because they point to "things out there", but referents because they direct attention to what we want to say something about.

So the difference between embodied and abstract concepts is this: while abstract concepts can be the formal referents in statements, only embodied concepts can be material referents. And the material referential relationship is one in which gestalts/engrams of physical objects (or features or processes) influence the choice of embodied concept words. Whether or not a currently perceived (or remembered, or imagined) physical object (via its gestalt) is matched to the LTM's engram of that object is a matter of degree; the match is a graded match, and the stronger the match is, the stronger the disposition to choose that word is.[31]

Of course, this is not all there is to choosing a word as the right one for forming a statement because all words also have Graded Meaning Sets which, as we have seen, are sets of first-word-to-second-word entailment- or inference-strength relationships, and also Generalization Sets which express our factual knowledge about what the concept refers to. So to the degree that a word is an abstract word – which is the degree to which perceptual gestalts/engrams do not influence its usage – it is the word's set of graded relationships with other words that influences the decision to choose or not choose it as the set associated with the word to be comprehended or the word to be produced.

This explains the experimental data that would seem to support the Context Availability Model over the Dual Coding Models. Direct lexicalized images are effectively missing as

30 I emphasized the importance of the set membership vs. set inclusion distinction in Part 1, where I explained what a concept is.

31 Note that the distinction between an in-awareness gestalt image, and an in-LTM engram image, and the matching of them, relies on a copy rather than a pointer metaphor of how sensory images and concepts are related to the memory of objects and processes. In lieu of that metaphor, which I am now inclined to deprecate, I would explain the "comparison" of an in-awareness to an in-memory image or concept as the activation of an in-LTM lexicalized image, or a concept, by the "site of control" STM narrative in nTM-narr having activated that particular LTM image or concept, by means of pointers. Nonetheless, I think there is something important to semantics in gestalts, and so I have left traces of this copy metaphor in this essay. The most important thing about gestalts, I think, is their all-at-once awareness. Without it, we could not understand the concepts we hear and speak, and the lexicalized images we perceive, without running through a list of all their intensional and referential relationships.

a constraint on abstract word choice. This is more consistent with the Context Availability Model, since what these contexts make available to the comprehension and production of statements is physical context, not as raw sensory input of course, but rather as lexicalized images.

The reason for the apparent difference between the two models is that, not distinguishing meaning from reference, the assumption is that embodied concept words have a provided context while abstract words do not. This isn't incorrect, but nTM explains what it means. For nTM, embodied concept words have a sensory-based referential force exerted on their selection whereas abstract words, if they do have such forces, have only weak ones, via intensional relationships to increasingly referential words. However, both embodied and abstract words have a co-ascriptional graded entailment- and inference- and generalization-set force exerted on their selection. So if words intensionally related to an about-to-be-comprehended or about-to-be-produced word have occurred in the set of statements already produced in the conversation or in the passage being read or composed (and are still, to some level of activation, in STM), that fact will encourage the selection of that word (for production) or the interpretation of that word (for comprehension).

I also point out that a strong referential force exerted on a word does not guarantee that it will be the word chosen. In concept-object referential associations, there is a concept, not just a named image! And that concept will have a Graded Meaning Set, not just a Graded Reference Set. If a word choice will entail the negation of many strong inferential relationships in that concept's Graded Meaning Set, that may be enough to cause us to "deny what is in front of our eyes", i.e. to deny that the referent object belongs to the kind indicated by the concept ascribed to it as the category it falls under.

Critique

The major shortcoming of the Dual Coding Models is that they recognize word-to-word associations, but cannot distinguish between associations due to meaning and associations due to reference. There is a statistical tendency for “horse” and “pasture” to co-occur in the same discourse contexts, and there is also a statistical tendency for “horse” and “animal” to co-occur in the same discourse contexts. The problem for the Dual Coding Models is this: there is no relationship of meaning between “horse” and “pasture”, but there is between “horse” and “animal”. “Horse” and “pasture” co-occur because horses are often found in pastures. This will be recorded in the Graded Generalization Sets in the lexicons. “Horse” and “animal” co-occur because being an animal is part of what it means to be a horse. This will be recorded in the Graded Meaning Sets in the lexicons. A horse is an animal *by definition*, and is found in pastures as a *matter of fact*.

Here is the word-world and dictionary-encyclopedia distinction that lexical semanticists have been searching for. Here is a semantic theory that accounts for the aboutness of language that truth-functional analytic philosophers emphasize, and that also accounts for the structuralism of language that post-Saussurean lexical semanticists emphasize.

Adding TM's distinction between meaning and reference sets to the Dual Coding Models allows those models to explain the processing differences between abstract and concrete words just as successfully as does the Context Availability Model, which is a model which reverts to the conceptual hemianopsia of meaning as reference which I alluded to earlier. nTM is far more neurologically detailed than are either of the Dual Coding Models. It accepts their prescient distinction of lexicons from the object recognition and integration areas. It explains the matter-of-meaning vs. matter-of-fact distinction within the lexicons, which the Dual Coding Models do not. And it explains the mutual influences of meaning and reference on one another in statement production and comprehension, by means of the Meaning/Reference Reciprocity Thesis. The MRRT provides a unified explanation of the neural basis for both concrete words and abstract words. I therefore propose nTM as a substitute for the Dual Coding Models and, of course, for the Context Availability Model.

nTM: Neuroscience Data Interpreted

The following items are only a small selection from the neuroscience empirical data about language and cognition which are reviewed in Kemmerer (2015) and in Banesh and Compton (2018).

nTM and the Interpretation of Aphasias

Here I will point out a difference between nTM and standard neuroscience theories with respect to the interpretation of some aphasia data.

Intra-ATL Aphasia

nTM hypothesizes that when connectivity between (i) Graded Meaning Sets and Graded Generalization Sets (the neural lexicon) and (ii) the object recognition and integration area is broken – i.e. when the two ATL hub pointers dissociate – a specific kind of aphasia will result. When the connection between meaning and reference is completely broken, but both the lexicons and the integrated object recognition areas are themselves unaffected, it will still be possible to identify objects as instances of their integrated images, i.e. to identify several pictures of horses as pictures "of the same thing", but it will not be possible to identify those objects as instances of their embodied concepts, or even as falling into the category indicated by the word's use as a semiotic sign.[32] And it will still possible to manifest lexical knowledge of the concept hitherto associated with that image. But it will be impossible to relate the two. We would not even be able to respond to a picture of a horse with the word "horse", or to pick out such a picture if given the word "horse". Both embodied concepts and semiotic signs would dissociate from the recognition of physical objects. We would no longer be able to sort the world into named kinds.[33]

This is not a conduction aphasia. It is an aphasia correlated with intra-ATL damage only. It is specifically the loss of a non-fascicular connection between the two pointers which make up an ATL hub (as nTM understands those hubs to be). Each of those pointers, of course, is a fascicular connection; but their association in the ATLs is not.

---

[32] Throughout this essay, I refer to integrated images of objects. This is a neurological-internal way of speaking. In all such cases, a more ontological way of speaking is also possible, one in which I would refer to objects, rather than to their neural images. That latter way of speaking is how we speak in our commonsense folk theory of the world and of our perception of it, and in most philosophical and linguistic contexts. Talk about objects belongs in Part 1, before I reached the neurological level. But here in Part 2, talk about images seems more appropriate, except when I think that the point is better expressed in the language of objects.

[33] However, this would not always be true, if nTM's hypothesis of non-ATL-mediated STM direct links between executive management and the object recognition and lexicon areas, is correct. What would fail would be only the *well-learned* paired links making up the hubs in the ATLs. But even with well-learned ATL-mediated links, direct links would be able, in some cases, to "pick up the slack".

Lexical Aphasias

Lexical aphasias are aphasias due solely to intra-lexicon damage. One form of lexical aphasia recognized in neuroscience is a breakdown in hyponym/hypernym relationships, such as that between "animal" and "horse". But since current neuroscience does not recognize semantic lexicons, with their networks of intensional relationships, they mis-interpret this data. An inability to traverse hyponymy trees, including those whose lowest levels are strongly associated with images, has nothing to do with images. It is caused by a breakdown, within the lexicons, of perhaps the strongest kind of entailment relationship. The subject cannot get from "horse" to "animal", for example, because that entailment-strength connection of the words in that ordered word-pair has been severed. A hyponym-named object is also a hypernym-named object because the latter concept is part of the meaning of the former concept. But with this kind of lexical damage, that knowledge is lost.

Note that it is not just damage to hyponymy relationships in the lexicons that manifests in impaired semantic behavior. All word-to-word impairments are lexical impairments. For example, consider a given NLU in the extreme case in which its wordform remains accessible in the lexicons but all of its connections to other wordforms are broken, but in which the ATL hub linking that wordform to an integrated image of an object is intact. In that case, conceptual knowledge could not be brought to bear on sensory experience, but semiotic proto-language signs could be. We could still respond with the word "horse" to a picture of a horse. In such a case, the ATL hub would still relate a word to an object, identifying the object as an instance of the kind denoted by the word. But that word would not bring along any additional conceptual knowledge with it.

The challenge for nTM then, as far as existing lesion/behavior research into aphasias is concerned, will be to see if the existing data can be re-interpreted so that it supports the hypothesis of neural lexicons, anchored in neural wordforms, as being a semantic functional area separate from the recognition of objects. It should also be possible to describe new lesion studies which would more precisely distinguish between the standard interpretation of such aphasia data and this nTM interpretation.

> Aside: Each year, I publish my notes as a "20xx Semantic Diary". 2016, 2017 and 2018 diaries are already published on my website tm1972.com, and at academia.edu. 2019 probably won't be published until mid-2020. I intend to include a nTM reinterpretation of a broader range of lesion/behavior data (and neural degeneration / behavior data) in that diary.

Word and Pseudoword Co-Equal Activation

> In Boatman et al's 2000 study, "essentially indistinguishable patterns of activation were found in the STS for words, pseudowords, and reversed words." (Kemmerer,

117-118).

Since nTM says that the phonological network is also the primary lexicon, how does it explain the fact that real words are not more highly activated than other phonetic sequences? The phonetic network / phonetic lexicon has several phonetic layers, leading from distinctive phonological features, through phonemes, to phonetic wordforms. Hebbian learning will create phonetic sequences which do not correspond to words because we encounter such sequences, e.g. when hearing two or more speakers speaking at the same time, or in the sequences in which the last phoneme in one word is followed by the first phoneme in the following word. We are immersed in a sea of learned phonetic sequences. Only some of them represent concepts.

Another idea is that as any sequence of phonemes is being heard, the brain begins a search process, trying to match an under-construction sequence, as each phoneme is added, to a sequence that represents a word. This is something like a spell checker which, at the most basic level, checks a word to be spell-checked against its dictionary one letter at a time, left-to-right. This would privilege wordforms over other phonetic sequences, of course.

Phonological and Semantic Density Neighborhoods

Kemmerer, at pp. 118ff, refers to a study by Okada and Hickok, 2006. That study found differences in phonological density within the phonological network and proposes this as an explanation of different levels of activity as words are chosen for production. But for nTM, the phonological network is also the phonological lexicon, and it contains not only sound-alike links between phonetic wordforms, but also semantic links – those described in Part 1 as making up the Graded Meaning Set and Graded Generalization Set for a word. So I propose that nTM-phonlex contains both phonological neighborhood density patterns and also *semantic neighborhood density* patterns. Just as “words from high-density (phonological) neighborhoods activate a greater range of phonological competitors than words from low-density neighborhoods”, words from high-density semantic neighborhoods will be found (if looked for) to activate a greater range of semantically-related words (not just competitors) than words from low-density neighborhoods.

> Aside: as I wrote this, I thought of the word “life” as an example. For most laypersons, the ordered word-pairs in which this is the first lexical item might include such second lexical items as “capable of growth”. For biologists, other strong or somewhat strong analytic (true by meaning) connections might include “sustained metabolism”, “capable of reproduction”, “adaptation to the environment”, and others.

If semantic neighborhood density is a fact, then neuroscience studies of activation in the neural lexicons (verbal or written) should show a more extensive spreading activation in the lexicon regions for words in semantically dense neighborhoods than for the others – in the example above for “life” in the brains of biologists compared to in the brains of laypersons. Eventually, neuroscience might even be able to locate the neural representations of individual wordforms and of their varying strengths of semantic connectivity with other wordforms.

Semantic Errors

> In an experiment by Hickok (2009), p.124, Kemmerer notes that “when patients make errors, they tend to be semantic in nature (selection of a semantically similar distractor picture) rather than a phonemic confusion (selection of a phonemically similar distractor picture”. (Kemmerer, 121).

(An example would be a picture of a moose selected as matching the word “bear”, rather than a picture of a pear.) This seems to be associated with bilateral lesions of the middle and posterior STG and white matter.

What is nTM's explanation for selecting a semantically-related rather than a phonologically-related picture? nTM says this takes place in the ATLs, whose pointers to the integrated object recognition area and to the lexicons conjoin meaning and reference. nTM's explanation of the semantic rather than phonemic substitution is then as follows:

The lexicalized image [Bear<--->[Bear]] is an ATL hub. The experiment involved a Wada procedure on the left hemisphere, and so the association of the word and the image was disrupted by the procedure. However, from nTM-narr (right>left), non-ATL-mediated direct links to integrated images and to embodied concepts can sustain a STM dynamic link between an image constellation and a word constellation relevant to the experimenter's question.[34] These constellations consist of bear-like images and of [Bear]-like concepts.

So “bear” exists in the neural lexicon in a semantic neighborhood where “moose” also exists. The word “pear”, although in the same *phonetic* neighborhood, was not in that same *semantic* neighborhood. The semantic neighborhood was the one selected, because semantics rather than phonetics was understood as most salient to the task asked of the subject. That semantic neighborhood, or a subset of it, was active as a constellation in nTM-narr. Because nTM-comm is not confined to one hemisphere, I speculate, it could survive the Wada procedure, and thereby enable the picture of a Moose to be selected as a match for the word “Bear” by narrative management.

[34] These constellations, note, because they are managed by nTM-narr, are on the contralateral side of the brain from nTM-stmnt, where a response to the experimenter's question will be formed. Being on the contralateral side, the Wada procedure did not affect them.

Word Deafness

> Some Wernicke's aphasics with unilateral left posterior STG/STS lesions are ... severely impaired at mapping subword-level phonological representations onto word-level phonological representations during speech perception tasks. (Kemmerer, 122).

nTM's explanation is that without word-pair semantic links, the phonetic sequences for words fade into the background of the phonetic network of the totality of phonetic sequences, making the ones for words difficult to pick out.

At (Kemmerer, 131ff), the Boatman (2000) study is referred to again. In this study, pMTG stimulation resulted in patients "hearing but not understanding the examiner" (Boatman, p.1631). In the primary lexicon, (partially in the pMTG) the links between phonemes which create the phonetic sequences of wordforms, can be disrupted. Semantic substitutions, e.g. "stick" for "pencil", reflect the perseverance of some semantic word-to-word links in the same semantic neighborhood, together with the wordform phonetic sequences for those semantic neighborhood words.

Left Bias for Phonemic Sounds

> Although all of the sounds … engaged areas in the dorsal STG bilaterally and to equal degrees, the phonemic stimuli engaged the middle STS in the left hemisphere to a significantly greater extent than the nonphonemic stimuli. (Kemmerer, 124).

So why would the left-hemisphere phonetic network be focused on those sounds more than on nonphomenic (nonmorphemic) sounds? nTM's explanation is that wordforms are phonemic (and thus phonetic) sequences, and the lexicon contains wordform pairs with links of different strengths. Given that wordforms are left-lateralized, it follows that the sounds that make them up will be the sound units that are co-located with them. These sounds are the ones we call phonemes, the phonetic units in wordform sound sequences.

This again emphasizes nTM's rejection of abstract wordforms as the loci of concepts. The loci of concepts are phonetic sequences, neurophysical wordforms, and nTM's hypothesis of this lexicon-located region for phonemic sounds is supported by the role it plays in the explanation of this and many other experimental results I have described. It is a better explanation than Kemmerer's point that "the phonemic stimuli were perceived categorically as familiar sounds, whereas the nonphonemic stimuli were perceived continuously as unfamiliar sounds". (Kemmerer, 125) For why are some phonetic stimuli – phonetic sequences and not just discrete phonetic units – more familiar? Because they occur in the spoken words we are constantly surrounded with, of course. But why would this create a wordform-supporting phonemic network, one apparently co-located with the perception of wordforms themselves? nTM's answer is that the two networks are both based on sound, with wordforms being sound sequences. And so it is the sound units found in those sequences which predominate in the left-hemisphere lexicon, nTM-phonlex.

Auditory Comprehension vs. Monitoring

At Kemmerer, p. 130, "cat" vs. "cot" is used to illustrate an impairment in which word-picture matching is intact while phoneme discrimination is impaired. "... Hickok and Poeppel point out (that) this double dissociation is 'remarkable'". But given the existence of a phonetic lexicon – a phonetic network which is also a semantic network – the phenomenon is not at all surprising.

If the intensional connections among wordforms in the phonetic lexicon are broken, then what is available via an ATL pointer will be a phonetic sequence only, the sound of the word. Relating words to pictures will be a monkey-like semiotic behavior, not a fully semantic one. In that case, however, the semantically disconnected wordform will still exist in a phonetic "sound-alike" neighborhood, but will not exist in a semantic neighborhood. Lacking any semantic constraints, phonetic drift is much more likely to happen; thus "cot" might easily be produced as a response to a picture of a cat.

Absence of Non-Initial Phonetic Substitution

How can we explain why phonetic substitutions do not regularly occur as non-initial sounds in words? The explanation I find supports in particular nTM's hypothesis of the co-location of a semantic network with the phonological network. I begin with the production of a phonetically correct wordform.

In writing or speaking, and with reading and listening as well, semantics is the task at hand. We are trying to express ourselves so we will be understood, or understand someone else who is trying to express herself. In speaking the next word in a sentence, for example, we are producing a sequence of phonetic units, each sequence being the sound of a word. Each word occurs in the semantic context of the other words one has produced, and so the selection of the next word to pronounce is a semantic task. Our brain isn't set to the task of selecting a word that sounds like the other words we have just spoken; it is set to the task of selecting a word that contributes to the meaning of the sentence we are producing.

With semantic objectives driving the speech we produce, and given a relevantly unimpaired brain, a best concept from a constellation of concepts will be chosen for the subject term of a statement to pick some thing out by saying what kind of thing it is, and a best concept from another constellation of concepts will be chosen for the predicate term to say something about it by pointing to a feature it instantiates. This leads to the selection of a phonetic sequence for a wordform in each case, but the task, to emphasize the point, is not to select a phonetic sequence; the task is to select a concept.

The process of speaking the desired concept begins with the first phonetic unit in that wordform, and proceeds to the next unit until the end of the word is reached. The first phonetic unit is selected because that concept, that locus in the lexical network, has been chosen as the right one for the semantic task at hand, and so semantically-driven selection picks out that wordform. Selection then proceeds to the next phonetic unit, guided by Hebbian-learned connections between one phonetic unit and the next one, and continues until the end of the wordform is reached.

So selecting the first sound in a word is governed purely by the semantic search in concept constellations in the lexicons for the semantically best word to use. But the selection of subsequent sounds is governed by an entirely different process. Frequently-used and frequently-encountered wordforms will be well-memorized phonetic sequences. Having one or more phonetic units in hand, the next one will naturally follow.

However, after an initial phoneme for a word is selected, there may be other less-well-remembered next-in-line phonetic units that lead to a non-word phonetic sequence. This is most likely to happen with the second phonetic unit selected, because once two phonemes have been selected, the number of remembered sequences which include one or more additional phonetic units, quickly narrows down. This first-to-last phoneme construction is analogous to a spell checker, which proceeds, orthographically, left-to-right. The further to the right a mis-spelling occurs, the more accurate the spell checker's suggested correction will be.

So the first phonetic unit for a desired word is selected under purely semantic constraints and so, unless the lexicon is severely damaged, it will be the first phoneme for the desired word. Successive phonemes are selected as remembered sequences, and even for fluent speakers, occasional non-words, or words other than the intended one, may occasionally be produced as successive phonetic units are selected – and, as I mentioned, these substitutions are more likely to occur the earlier in the sequence the phonetic unit is.

A more detailed analysis will lead to additional insights into nTM. The learned phonetic sequence for a word makes the selection of a second syllable easy, since it is the second sound element in a small set of semantically relevant words, rather than in the much larger set of all the words in the lexicon, or the even larger set of all the phonetic sequences in the co-located phonetic network. By the same token, after two phonemes have been assembled in WM/STM, the set of correct words may already be at a single word, or if not, at a small number of them. As the sequence is assembled, subsequent phonemes quickly narrow down to the one desired word. All of this is simple Hebbian learning of the phonetic sequences for words.

But this set of links among adjacent phonemes can be broken by lesions or neural degeneration in the phonetic lexicon, so that non-initial phonetic substitutions may occur.

Perhaps the fact that they rarely occur indicates that the phonetic sequences making up wordforms are extremely well-learned in the phonetic lexicon, relative to non-wordform sequences. If so, we might find that non-initial phonetic substitution is relatively more common with children, and with others who have been less immersed in the sea of language than most of us.

On the Dual Stream Combinatorial Network for Meaning

At pp. 134ff, Kemmerer discusses the Hub and Spoke Model's **c**ombinatorial network for meaning. The model distinguishes a left lateral ATL area for syntax, and another area for compositional semantics. (References to Vandenberghe et al (2002); Humphries et al, (2006); Rogalsky and Hickok, (2009)). The last experiment discussed distinguishes passive listening vs. attention for semantic and syntactic anomalies in sentences. There were four related conclusions, each of which I discuss in the rest of this section.

1. Stronger response in entire ATL ROI for attention vs. passive listening.
2. Most of ATL equally engaged for semantic and syntactic anomalies.
3. Some of ATL may be more sensitive to semantic than syntactic anomalies.
4. No area more sensitive to syntactic than semantic anomalies.

<u>The first conclusio</u>n reflects the fact that in listening to others in conversation, we are very tolerant of semantic and/or syntactic errors. If we were not, conversations would be interspersed with each of us correcting the other, or asking for clarifications of what he meant by certain lexical items or sentences. So we work hard to figure out for ourselves a narrative understanding of what the speaker is saying as semantic and syntactic anomalies occur in his speech. We “cut him a little slack”, for the duration of a conversation, by allowing references we normally would not allow, and co-ascriptions we normally would not allow. To continue to make sense of a narrative containing non-standard semantics for both embodied and abstract concepts, and also syntactic errors (making it harder to keep track of who did what to whom) – this requires hard work. Temporarily-accepted extensions to the meaning and/or reference of concepts will conflict with the established connections existing in the lexicons, and also in the ATL hubs.

nTM predicts that a similar strong response will also be occurring in the lexicons. It is deviation from the standard meanings of words, not just from their standard references, that must be tolerated and interpreted.

<u>The second conclusion</u>. In STM/WM, the statement construction area constructs sentences one lexical item at a time. That area contains two concept constellations, provided by the narrative management area, reflecting the intention to pick some thing out by saying what kind of thing it is, and the intention to say something about it by picking out a feature of that referent that we have noticed. So in the statement construction area, the task is not simply the syntactic task of linearizing an already-constructed phrase-structure diagram, with lexical nodes already filled in. If that were the

case, there would be no need for semantic processing in nTM-stmnt.

Instead, what happens is this: two constellations of concepts already exist in the nTM-exec area, localized to nTM-narr. Before we produce a sentence, we know what we want to talk about – the referent of the sentence – and we know what we want to say about it – the feature we wish to direct our listener's attention to. The statement construction area already has its subject and predicate components roughly sketched out, as the currently most active concept constellations in narrative management.[35]

Since the distinction between subject and predicate is a syntactic one (as well as a semantic one), some neural activity we associate with syntactic processing takes place in nTM-stmnt. But, as I said, specific lexical items have not at that point been picked out. The semantic work of managing concept constellations dominates the activity of nTM-exec. This semantic work, of course, involves (for concrete words) finding a semantic group of concepts that cluster around the perceptual image the speaker is focused on; the clustering is one subject to matching the current image with LTM images, and finding a good enough match to "pass muster".[36] For all words, this semantic work also involves finding wordforms whose analytic, synthetic and generalization word-pairs do not support attempts to discredit what the speaker wants to say as a mistake.

I have already (in Part 1) described the two ways that a basic statement can be rejected. One is to claim that the referent and/or its denoted feature, are not correctly picked out by the embodied concepts used. Perhaps I have used the word "clump of sand" to denote what the other party can successfully argue is obviously a rock. The other is to claim that the referent and/or its denoted feature violate consequences of the meanings of the concepts, e.g. that the referent can't be an A because all As are B and all Bs are C, but the referent is not a C, e.g. that all horses are animals and all animals are made up of cells, but that object is made up of minerals, not cells. Again, temporarily-accepted extensions will conflict with the established word-object hubs in the ATLs.[37]

[35] This is not a "knowing about comes first" mentalism. The "knowing about" is something to be explained neurophysiologically. It is the linguistic manifestation of the "executive planning and control" function which is being actively investigated in current neuroscience. I look at executive planning and control as a form of pattern recognition, which the brain is very good at, including the projection of perceived patterns into the future, i.e. perceiving them atemporally, and sometimes not the most dominant perceived pattern. Where language is concerned, concept constellations and statements are the basic elements in these patterns.

[36] This is a reversion to the copy metaphor. An equivalent description, in terms of the pointer metaphor – one which emphasizes the distinction between sources of control and sites of activity – is easily provided.

[37] As in Magritte's famous "Ceci n'est pas une pipe" painting of a pipe, if my interlocutor wants to refer to a statue of a horse as a horse, I'll be ok with that. But this is an extension of meaning, one not consistent with the well-established lexicalized image / embodied concept [pipe] (or [horse]).

Finally, in accordance with TM's MRRT, the final selection of a specific lexical concept, done in the statement construction area, attempts to find a reconciliation between the perceivable features of the referent and its noted feature, and the meaning sets for the concepts. This is why the area of the brain in which sentences are constructed involves semantic work and not just the syntactic process of linearizing phrase-structure trees.

Going back to a point I made in Part 1, these two ways of challenging a basic statement correspond to the two components of mathematical sets, which I there extended to include dynamic sets. One way is to challenge the extension of one or both concepts. What they refer to, it is then said, does not belong as a member of the set denoted by the concept. The other way is to challenge the set membership criterion for the set. The set membership criterion is the Graded Meaning Set for the concept, and at its core are the analytic statements in which the concepts are the first concepts in ordered concept pairs.

<u>The third conclusion</u>. nTM-narr does not involve phrase-structure construction or its linearization into produced sentences. Its work is to activate, in the lexicons, concept constellations for the subject and predicate of the to-be-produced statement. This involves activating engram images in the LTM nTM-what and nTM-where areas, via the integrated object image pointers in the ATL hubs or directly via the nTM-OccipFrontFac, subject to the Meaning Set constraints that the concepts for the to-be-produced co-ascription do not contain entailments or inferences or well-established generalizations inconsistent with recently-produced statements or with one another. This produces two constellations of concepts – rough sketches of subsets of wordforms from which nTM-stmnt will select the specific concepts to use.

<u>The fourth conclusion</u>. As already pointed out, nTM-stmnt does not do purely syntactic work. Its semantic work is to select specific concepts, one from each constellation whose heightened activation is due to the work of nTM-narr. The work of nTM-narr will engage temporarily-accepted extensions to either or both of the images or concepts related as hubs in the ATL, but those extensions themselves, and their conflicts with the LTM established ATL hubs, are not syntactic ones.

Picture Naming vs. Word Comprehension

> The Boatman et al (2000) study "demonstrated that direct cortical stimulation at some sites in the left mid / posterior MTG does not affect picture naming (suggesting preservation of the output mapping from semantics to phonology) but nevertheless impairs word comprehension (suggesting disruption of the input mapping from phonology to semantics". (Kemmerer, 159).

For nTM, there is neither "output mapping from semantics to phonology" nor "input mapping from phonology to semantics". Semantic and phonological links co-exist in the phonological lexicon. So what does the Boatman data indicate?

From the nTM perspective, it is no surprise that cortical stimulation of the lexicon areas may often leave picture naming unaffected. What disruption of the lexicons does, in these instances, is cut off the lexicons from the ATL processes of naming kinds of things with concepts. But it may leave behind the ability to name those kinds of things with semiotic signs, with the same resources that monkeys have with their proto-language abilities. This implies that the intra-hub association of an image and a wordform, in the ATLs, is still operative, which in turn suggests that the pointer from an ATL hub lexicalized image to its lexicalizing wordform in the lexicon, is also still operative. What has been damaged are the intra-lexicon intensional relationships of that word to other words in the lexicon.

And, of course, it is no surprise that word comprehension is affected. The lexicons are where the intensional/meaning components of the semantics of words are stored in LTM. Wernicke was not far off.

> Kemmerer also notes that "… when some patients perform various word production tasks, they make semantic substitution errors in spoken output but not in written output, whereas other patients exhibit the opposite dissociation." (Kemmerer, 166)

nTM has a straightforward explanation of this dissociation. In the one case, it is the phonetic lexicon that is damaged; the desired word can't be reached, so nTM-stmnt, in WM, searches for a semantic nearest neighbor. Since the intensional component of semantics exists within the lexicons, there is no puzzle to solve about where, along the path from meaning to sound or shape, a mistake occurred. And the same explanation applies for the orthographic lexicon. The two lexicons are dynamically synchronized; but they can never be so coordinated as to be two modalities of an identical set of entailment, inference, and generalization relationships.

Pictures and Words in the Fusiform Gyrus

> As the authors point out, Kemmerer notes, (Chao et al, 1999) "it is especially interesting that these adjacent but nevertheless distinct regions of the fusiform gyrus were activated not only by pictures, but also by words". (Kemmerer, 279)

There are two possible explanations available to nTM. One relies on my interpretation of ATL hubs as lexicalized images / embodied concepts. On this account, one of the two coordinated pointers in an ATL hub connects the hub to the integrated image, while the other coordinated pointer connects the hub to the wordform which lexicalizes that image. The result is a lexicalized image / embodied concept, which is the hub itself. In this case, activation of the fusiform / posterior-occipital /inferior temporal lobe area by the name for a kind of object take place because of the activation of the corresponding ATL hub.

I don't know whether or not neuroscience data supports this hypothesis. If it does not, that implies that there may be a direct lexicon/fusiform connection. If we assume that a monkey's “lexicon” is simply a collection of isolated phonetic semiotic signs – which is what current neuroscience takes concepts to be – then the monkey's use of a phonetic sign for tigers could be supported without anything analogous to an ATL hub. Of course, as already mentioned, an alternative explanation of the preservation of semiotic signs when concepts are no longer accessible is that the intensional relationships for a specific wordform, in the lexicons, have been damaged.

nTM: One Last Time

Unlike other neuroscience semantic models, as Part 1 makes clear, nTM is a theory with deep roots in analytic philosophy of language and in lexical semantics, and also brings together long-standing truth-functional and structuralist approaches to the study of language. There is much that better-directed neuroscience experimental research could do to contribute to our understanding of semantics, and my hope is that nTM will provide that direction.

The basic neural components of TM are, of course, those that correspond to the reference and meaning of concepts, and to their ascriptive and co-ascriptive uses in statements.

At the philosophy and lexical semantics levels, *meaning* is a variable-strength co-ascriptional association between ordered pairs of concepts, expressed as universal quantifications which are entailments and inferences, periodically recorded in dictionaries as definitions. At the cognitive psychology level, it is a variable-strength stimulus-response disposition to use or to accept the use of a second word in an assertion in which a first word has already been used. At the neural level, it is a variable-strength association between two neural groups, each being one which fires when a wordform is perceived, and one of which is available to the working memory task of assembling a statement because the other is already in place.

At the philosophy and lexical semantics levels, *reference* is a variable-strength ascription relationship between a concept and an object, giving a circumscribed hope to correspondence theories of truth. At the cognitive psychology level, it is a variable-strength stimulus-response disposition to use or accept the use of a word to say what kind of thing some perceived image represents, or what kind of feature it instantiates. At the neural level, it is a variable-strength association between two neural groups, one tending to fire when a wordform is recognized, thus bringing the associated image into awareness, and the other tending to fire when an object, feature, or process is recognized, thus bringing the associated word into awareness.

At the philosophy and lexical semantics levels, a *basic statement* is a co-ascription of two concepts to a referent which is an instance of some kind, and which instantiates a noticed property or relationship which is also an instance of some kind. When the speaker or listener believes that the subject term referent is of the indicated kind, then the ascription of the subject term is accepted, and similarly for the ascription of the predicate term. The Graded Reference Sets of the two concepts have done their work. It is then, literally, a matter of whether or not we can believe what is in front of our eyes.

If there are no inter-conceptual links to either of the concepts which are sufficiently strong to raise the possibility that their co-ascription in the statement may violate intensional constraints created by those inter-conceptual linkages, then we can believe what is right in front of our eyes. If we can't, then for either or both of those concepts – and in spite of each satisfying its own referential constraints – something is wrong. What is wrong can always be proposed like this: "You say that the [A] thing over there is also a [B], and indeed so it appears to be. But if it is a [B], then it is (definitely or possibly) a [C]" (an assertion which can be an entailment, an inference or a generalization) "and we both know that no (or only a few) [A] are [C], so that [A] over there can't be (or is unlikely to be) [B]".

The resolution must alter, for those persons on those occasions, either reference or meaning or both, either concept-to-object links or concept-to-concept links. Perhaps Graded Meaning Sets will dominate in this instance of resolution. The object in question may have looked only marginally like an [A], or its noted feature only marginally like a [B]. In that case, we are likely to henceforth exclude that object or feature from the referential range of [A] or [B], to avoid intensional tensions.

Doing this alters the neural links for the referential use of [A] and/or [B]. The reference of [A] and/or [B], for you and I, on subsequent occasions, will have changed; and when that happens for enough people, on enough occasions, the exclusion of things very much like the thing designated on that occasion as an [A], or the feature designated on that occasion as a [B], will spread via exemplification to enough other people, on enough occasions, that eventually it will become correct to say that the linguistic community itself has accepted that constriction of the Graded Reference Set of either or both of those concepts.

On the other hand, perhaps the Graded Reference Sets will dominate in this instance of resolution. The object in question may have been a prototypical instance of [A], and its noted feature of [B]. If we were hesitant to accept that the [A] in question was also [B] because we believed that all [B] are [C] and also that no [A] are [C], then we may decide that the [A] in question is actually a counterexample, showing us an [A] which indeed is [C] and/or a [B] which is not [C]. In that case, either or both of those inferences, or perhaps even entailments, will have to be demoted to generalizations. Or if we were hesitant to accept that the [A] in question was also [B] because we believed that most [B] are [C] and/or that most [A] are not [C], then we may decide that we should demote either or both of these generalizations, say from "most" to "some", or from "some" to "a few".

Meaning and reference constitute the semantics of concepts whose co-ascriptions and ascriptions in various types of statements express our beliefs about the world. Via cognitive sociological processes, the semantic networks in each of our brains continually adapt to one another to produce a dynamic convergence which allows us to share our knowledge, and to record it in our evolving dictionaries and encyclopedias.

Appendix: Exporting nTM to Silicon

As a set of functions and their interrelationships, the nTM Semantic Model can serve as a high-level architecture for a language producing and language comprehending machine. One step to starting up the machine would be to load it with a reasonably robust list of co-ascribed word-pairs, each with a selection of entailment, inference and generalization associations and, for embodied concepts, of word-to-image associations. This is what we do when we teach language to our children.

For a machine to make a basic material-referent statement would be for it to produce (i) an image of an object, (ii) a word representing its lexicalizing concept, designating what kind of thing the image represented, (iii) an image representing a noticed feature of that object, and (iv) another word designating the kind which that feature instantiates. Another machine, comprehending the statement, might accept it, or perhaps might reject it as co-ascribing a word together with the negation of another word to which it is strongly related by meaning. Or it might less strongly reject it, accepting it as a possibility but one which does not happen to be instantiated in the real world. In these processes of trying to understand one another, each machine might adjust the penumbral ranges of its lexicalized images, and the strength of its co-ascriptional relationships with other words. It might do all of the things we do in producing and comprehending language; and the Graded Meaning Sets and Graded Reference Sets of those machines would then evolve in similar ways to the way that our sets evolve. Graded Generalization Sets would also form and evolve, recording the increasing knowledge the machine would have about the world around it.

Pattern recognition is something cANNs (convolutional artificial neural networks) are getting good at. But the trouble with neural AI proficiency in matching a current image – of a handwritten letter of the alphabet, in early neural AI work – is that it is difficult to understand *how* an ANN achieves its increasingly amazing pattern-matching results. From a nTM perspective, it might go something like this.

1. The initial layer or layers of the ANN record the input image. The initial layer will be a pure transducer, without any interpretative image-manipulative work going on. Subsequent early layers will produce more complex consolidations of that basic input, along the model of the visual and auditory processing streams in the human brain.

   These subsequent early layers correspond to the brain's short-term memory of a present perceptual experience. And in the brain's case, we know a great deal about the image-manipulative work going on after that and before we reach awareness of the image as an image of some kind or other of thing, particularly in the case of visual imagery. Something similar may be hypothesized to be going on in the early layers of a cANN.

2. Middle layers must already contain LTM consolidated images. Those images already exist because with each input that results in the correct identification of a letter, the input image is back-propagated to modify the consolidated image it was matched to.

   This corresponds to the process in which we learn, by trial and error, which embodied concepts are referentially associated with each of the images they lexicalize, and the Hebbian processes by which consolidated lexicalized images are stored, modified and accessed by LTM pointers in the ATL hubs.

3. In those middle layers, the STM input pattern is matched to a range of LTM consolidated images, those that constitute possible matches. Later, near the output layer, one of those images is selected as the best match. The output layer displays that image as the result of the match process. Of course, much more will be learned about how neural networks (human and silicon ones) select best matches, but there is also a great deal which neuroscience knows about the process already.

4. If the result is judged correct – by the community of language users (or, eventually, of other machines) judging those results – the image is then back-propagated to the consolidated image for that letter. I note that this is very much like the process in which language users learn the correct referential use of their embodied concepts by adjusting their referential behavior when corrected by other fluent speakers.

The LTM images exist along a continuum of perceptual similarity which the ANN (in these early experiments) is constrained to divide into one of twenty-six categories, or perhaps into one of fifty-two. This continuum is not smoothly graded. It is bumpy. Each letter's consolidated image is one of those bumps, a local minimum on a contour map.

How does an ANN do increasingly well at correctly identifying inputs? It is approximately the same way that monkeys do increasingly well at making tiger calls when a tiger is present, and different calls when different sources of danger are present.

The unsuccessful monkeys will not survive to pass on genes coding for brains that make such mistakes. Fortunately, we are more forgiving of our machines, and of our own language learners. But here is an important point. If the MRRT is correct, then since language users refer to things with *concepts*, not just with signs, incorrect references can be identified, even when closer inspection doesn't change the images, if those references entail a conflict in the meaning of closely-associated concepts. That thing out there may be perceptually indistinguishable from a tiger, but if it turns out to be a virtual reality projection of a tiger, which entails among other things that it is not a real animal, then it isn't a tiger, and our use of the concept [Tiger] – as the wordform "tiger" – to refer to it will stand as an error – an understandable one, but an error nonetheless.

Based on a superficial understanding of neural net AI, I suspect that the pattern-matching prowess of these machines mimics the monkeys' use of *signs* to label kinds of things, but not the human use of *concepts* to label them. In the brain, image formation from scanning the surface of our visual transducers produces initial input which is then processed, in both dorsal and ventral streams, to eventuate in both the recognition of discrete objects, and in the awareness of their relative location with respect to other objects, in both space and time. This process is facilitated by biologically hardwired responses to initially very simple visual patterns, and also to increasingly complex visual patterns until, in anterior areas of the parietal, temporal and frontal cortices, *learned* pattern recognition lights up. I propose that the first learned patterns correspond to the visual objects most important for biological survival – predators, prey, competitors and mates.

Convolutional neural networks are more closely modeled on the human brain than classical neural networks are, in two ways. First, convolutional neural networks clump the processing of visual input into stages, just as the brain does as input proceeds from the occipital cortex forward through the V1-V5 areas primarily located in the parietal cortex. In the brain, these stages are functionally identified by the type of visual input they respond to, and by the results they make available to downstream visual processes. These input-processing-output clumps are neurally hardwired, in response to evolutionary forces we can understand in only the broadest terms.

Second, convolutional neural networks are not fully connected between layers. Connections between pairs of elements in adjacent nodes will be dropped if they fail to contribute to a match for a long-enough period of time. This reflects, precisely, the complement of Hebbian learning – Hebbian anti-learning. Neurons that fire together wire together; and neurons that don't fire together loose any wiring they may have had.

If our objective is to create silicon intelligence, then copying the human brain isn't a bad idea. We will be copying an inelegant implementation of intelligence, however, because Nature is, of necessity, opportunistic. It solves current problems by adapting what it has to hand, and has no way to periodically reconceive the entire architecture, starting over again from a clean whiteboard. But, the point is, it works. The brain does enable intelligent behavior. So in the visual input stream – and other sensory modalities as well – we have a clumping of processes leading to an end result which can enter conscious awareness.

If so, the next step in the evolution of neural net AI should be to generalize from co-ascriptional occurrences of word-pair patterns in basic statements, and to record these generalizations either as generalizations (in TM's technical sense), or as inferences, or else as entailments. With those patterns encoded, constantly modified, and used in the production of basic statements – perhaps in the syntactically-simplified form of indexicalized ordered and unordered word pairs – we would be well on our way to language-mediated intelligence implemented in silicon. Machines would be able to produce language, to comprehend it, and to continually adjust their own semantic networks to those of other machines and, hopefully to those of we human beings too, resulting in the ability to share information.

We should now have a pretty good idea of how Graded Sets are created, modified and accessed in the human brain, and of what could be done to introduce Graded Meaning Sets and Generalization Sets to neural network AI machines. I have introduced this comparison to ANNs because it has allowed me to present the use of concepts in a way which abstracts from specifically human-brain neural implementations, and to once again emphasize that having names for kinds of things is not the same thing as having concepts for them.

The architecture of ANNs is an architecture of layers of nodes, each node connected to any number of nodes in an adjacent layer or in its own layer. The human brain has evolved, ultimately, from a single interneuron connecting a sensory input neuron and a motor output neuron.

The question I have is this: did the brain develop a complex architecture of functional neural areas and connections between them simply under the pressures of speeding up signaling as animals began their predator/prey co-evolution? That is, except for speed, and perhaps also economy of the use of energy within the system, could the human brain do what it does now if all it had was an ANN-like architecture?

It would seem so. And if so, ANN engineers probably don't have to consider discarding the architecture of layers in their ANNs. This is not to say that ANN engineers have nothing to learn from neuroscience. For example, they have relied heavily on what neuroscientists have learned about the human visual system in constructing robotic cANN-driven visual systems. But the main architectural feature of the brain's visual system – as indeed it is for all the brain's sensory input processing – is a step-by-step progression from initial transduction, through very basic hardwired processing, to increasingly complex intermediate constructs for which the wiring implements learned patterns, and finally, to an end product in sensory awareness to which an intelligent response can be made.

Is this not a layered architecture, with backpropagation? If so, is it perhaps only in the wetware vs. silicon details that there is any substantial difference?

Evolution always produces, at each point in time, the best it can make, by means of natural selection, of the previous state of the environment, the organism, the organ, the cell, the molecule, etc. And in spite of rule-governed copying, random changes sometimes occur in the copying process. Novelty is introduced. Without the novelty introduced by the death of sentient entities and the production of essentially similar new sentient entities, silicon intelligence would remain static. Immortal silicon machines would have difficulty surviving in an ever-changing universe.

The point to notice here, once again, is that Nature does not begin with a blank slate, a clean whiteboard. It works with the latest result of the best it can do with a product that conforms, for the most part, to a previous pattern, but that also includes random modifications. Since that is the origin of the human brain, why should silicon engineers attempt to reproduce such a product in all its intricate architectural detail? Why should they not learn what they can from Nature, and then, in mimicking Nature, attempt to start afresh and design the cleanest architecture they can?

A Science Fantasy Thought Experiment

I note also that implementing these clumps in silicon isn't arbitrary, nor is it cheating, in some way, in the task of creating silicon intelligence. My reasoning is this: if non-biological intelligence has evolved elsewhere in the universe, perhaps it too was jump-started by a species of biological engineers. Perhaps it *had* to be that way.

Here's why. All too briefly, without predator/prey relationships, the only thing that could motivate an organized group of matter (like an embodied neural network) to do anything would be an awareness of the consequences of the second law of thermodynamics, in all its infinite variations, and the desire to continue to exist, or to produce copies that would continue the existence of the pattern it instantiates.

But that second option won't do because to be aware of the second law of thermodynamics, it would have to exist in the first place. So we are back to the first option – the need to consume energy and acquire raw material in the context of an ecosystem of predators, prey, competitors and mates. Without a pre-existing desire to do that, we are left with no explanation of how biochemical processes taking place within a container that regulates the inputs and outputs of the entity it is, could arise in the first place. To get started, to avoid the chicken and the egg problem, something like the earliest cells – groups of molecules in a selectively permeable membrane that both defined the distinction between self and other, and also protected the dynamic evolving molecular structures within, and could also bud off copies of themselves – had to arise by accident. But, as quantum physicists often say, anything that could happen in the universe eventually will happen.

The incessant activity of matter, at the atomic level, was the raw material for those accidents to occur. In order to be complex enough to carry out processes of transforming inputs into the material needed for structural growth, repair and reproduction, and in the energy needed for those transformations, carbon and water were essential – carbon as the essential material component, and water as the essential catalytic component. As these groups produced copies of themselves, local environments would begin to see competition for resources, and also activity in which one group could be the material input for another group.

A basic constraint on the explosion of life is the scarcity of matter and energy in a form that an organism is able to utilize. This constraint alone would produce an incentive to evolve. But a more advanced constraint is the predator/prey arms race. Once one type of cell became able to utilize another type of cell as input, the first type became a predator, and the second its prey. Ecology tells us that predators cannot exist without prey, and so an arms race became a permanent feature of organic life, even the arms race among plants and between plants and animals.

Occasionally, ingestion of a second type of cell does not result in the breakdown of the second type of cell into raw molecular material. In this way, animal cells acquired their mitochondria, and also their ingested bacteria and viruses.

So I propose that, first, intra-stellar pressures led to the full array of stable atoms we find in the universe. Next, random collisions among atoms, in cooler extra-stellar environments, led to the formation of carbon and water planets on which the first complex, organic molecules could evolve. Next, the combinatorial possibilities of carbon, hydrogen and oxygen, and the catalytic ability of water to facilitate breakdowns and recombinations of these molecules, made it possible for the first encapsulated collections of organic molecules to appear. Eventually, among the many different types of these encapsulated collections were ones that were able to divide and reproduce. As biochemical processes, these processes needed matter and energy inputs to sustain themselves. They were the necessary foundation for any form of intelligence in the universe – including silicon intelligence.

In the course of the evolution of intelligence in the universe, perhaps intelligent wetware creatures such as ourselves are a unique point in the history of that evolution – the point where the vicissitudes of organic evolution can be left behind, and a more elegant, beautiful, efficient and successful form of intelligence come into being.

Bibliography

Banesh, Marie T. & Compton, Rebecca A. *Cognitive Neuroscience, 5th edition*. Cambridge University Press, 2018.

Bickerton, Derek. *Adam's Tongue*. Hill and Wang, 2009.

Bickerton, Derek. *More than Nature Needs*. Harvard University Press, 2014.

Binney, R.J., Embleton, K.V., Jeffries, E., Parker, G.J M. & Lambon Ralph, M.A. “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.” *Cerebral Cortex*, 20 (2010), 2728-2738.

Boatman, D., Gordon, B., Hart, B., Selnes, O., Miglioretti, D. & Lenz, F. “Transcortical sensory aphasia: Revisited and revised.” *Brain*, 123 (2000), 1634-1642.

Bright, P., Moss, H.E., Stamatakis, E.A. & Tyler, L.K. “Longitudinal studies of semantic dementia: The relationship between structural and functional changes over time.” *Neuropsychologia*, 46 (2008), 2177-2188.

Brown, C.M., & Hagoort, P. (eds.), *The Neurocognition of Language*. Oxford University Press, 1999.

Carnap, Rudolph. *Der Logische Aufbau der Welt*. Felix Meiner Verlag. 1928. Translated as *The Logical Syntax of Language*. Harcourt Brace, and Kegan Paul, 1937.

Chao, L.L., Haxby, J.V., & Martin, A. “Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects.” *Nature Neuroscience*, 2 (1999), 913-919.

Damasio, A.R. “Concepts in the brain.” *Mind and Language*, 4 (1989), 24-28.

Damasio, A.R. *The Feeling of What Happens*. Harcourt, Brace & Co, 1999.

Dehaene, Stanislas. *Consciousness and the Brain*. Penguin Books, 2014.

Eco, Umberto & Santambrogio, Marco & Violo, Patrizia (eds). *Meaning and Mental Representations*. Indiana University Press, 1988.

Gazzaniga, M. (ed.), *The New Cognitive Neurosciences*. MIT Press, 2000.

Gazzaniga, M.S. (ed.). *The Cognitive Neurosciences, 4th edition*. MIT Press, 2009.

Harris, Randy Allen. *The Linguistics Wars*. Oxford University Press, 1993.

Hauser, Larry. “Chinese Room Argument”. *Internet Encyclopedia of Philosophy*, 2018.

Hickok, G. “The cortical organization of phonological processing.” In [Gazzaniga, M.S. (ed.), 2009a], 767-776.

Hickok, G. “The functional neuroanatomy of language.” *Physics of Life Reviews*, 6 (2009), 121-143.

Hickok, G. & Poeppel, D. "The Cortical Organization of Speech Processing." *Nature Reviews Neuroscience*, 8 (2007), 393-402.

Hoffman, P. & Lambon Ralph, M.A. "Reverse Concreteness Effects Are Not a Typical Feature of Semantic Dementia. Evidence for the Hub-and-Spoke Model of Conceptual Representation". *Cerebral Cortex*, 24 (2011), 1767-1777.

Humphries, C., Binder, J.R., Medler, D.A. & Liebenthal, E. "Syntactic and semantic modulation of neural activity during auditory sentence comprehension." *Journal of Cognitive Neuroscience*, 18 (2006), 665-679.

Indefrey, P., & Levelt, W.J.M. "The neural correlates of language production." In [Gazzaniga, M. (ed.), 2000], 845-865.

Indefrey, P., & Levelt, W.J.M. "The spatial and temporal signatures of word production components." *Cognition*, 92 (2004), 101-144.

Kemmerer, David. *Cognitive Neuroscience of Language*. Psychology Press, 2015.

Kousta, S.T., Vigiolcco, G., Vinson, D.P., Andrews, M., & Del Campo, E. "The representation of abstract words: Why emotion matters." *Journal of Experimental Psychology: General*, (2011), 14-34.

Lambon Ralph, M.A. & Patterson, K. "Generalization and differentiation in semantic memory." *Annals of the New York Academy of Sciences*, 1124 (2008), 61-76.

Lenci, Alessandro. "Distributional Semantics in Linguistic and Cognitive Research". *Rivista di Linguistica* 20.1 (2008), pp. 1-31.

Lenci, Alessandro. "Will Distributional Semantics Ever Become Semantic?" *7th International Global Wordnet Conference*, January 28, 2014 – Tartu.

Levelt, W.J.M. *Speaking: From Intention to Articulation*. MIT Press, 1989.

Levelt, W.J.M. "Models of word production." *Trends in Cognitive Sciences*, 3 (1999), 223-232.

Levelt, W.J.M. "Producing spoken language: A blueprint of the speaker." In [Brown, C.M. & Hagoort, P. (eds.), 1999], 83-122.

Levelt, W.J.M. "Spoken word production: a theory of lexical access." *Proceedings of the National Academy of Sciences*, 98 (2001),13464-13471.

Lyons, John. *Linguistic Semantics*. Cambridge University Press, 1995.

McClelland, J.L., Rogers, T.T., Patterson, K., Dilkina, K. & Lambon Ralph, M. "Semantic cognition: Its nature, its development, and its neural basis." In [Gazzinga (ed.), 2009], 1047-1066.

Paivio, A. *Mental Representations: a Dual Coding Approach*. Oxford University Press, 1986.

Paivio, Allan. *Mind and Its Evolution: A Dual-Coding Theoretical Approach*. Lawrence

Erlbaum, 2007.

Paivio, A. “Dual coding theory: Retrospect and current status.” *Canadian Journal of Psychology*, 45 (1991), 255-287.

Patterson, K., Nestor, P.J. & Rogers, T.T. “Where do you know what you know? The representation of semantic knowledge in the brain.” *Nature Reviews, Neuroscience*, 8 (2007), 976-987.

Petersen, S.E. et al. “Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing”. *Nature*, 331 (1988), 585-589.

Pulvermuller, Friedmann. *The Neuroscience of Language: On Brain Circuits of Words and Serial Order*. Cambridge University Press, 2002.

Quine, W. V. "Two Dogmas of Empiricism". In [Quine, 1961], 20-46.

Quine, W. V. *From a Logical Point of View*. Harper Torchbooks, 1961.

Quine, W. V. *Word and Object*. MIT Press, 1960.

Rogers, T.T., Lambon Ralph, M.A., Garrard, P., Bozeat, S., McClelland, J.L., Hodges, J.R. & Patterson, K. “The structure and deterioration of semantic memory: A neuropsychological and computational investigation.” *Psychological Review*, 111 (2004), 205-235.

Rosch, Eleanor. “Cognitive Representation of Semantic Categories”. *Journal of Experimental Psychology: General*, 104 (1975), 192–233.

Santambroglio, Marco & Violi, Patrizia. "Introduction to Meaning and Mental Representation". In [Eco, et. al. 1988], pp. 3-22.

Searle, John. "Minds, Brains, and Programs." *Behavioral and Brain Sciences* 3 (1980), 417-424.

Simmons, W.K. & Barsalou, L.W. “The similarity-in-topography principle: Reconciling theories of conceptual deficits.” *Cognitive Neuropsychology*, 20 (2003), 451-486.

Vandenberghe, R., Nobre, A.C. & Price, C.J. “The response of left temporal cortex to sentences.” *Journal of Cognitive Neuroscience*, 14 (2002), 550-560.

Vigliocco, G., Kousta, S., Della Rosa, P.A., Vinson, D.P., Tettamanti, M., Devlin, J.T. & Cappa, S.F. “The neural representation of abstract words: The role of emotion.” *Cerebral Cortex*, 24 (1997), 1767-1777.

Vigliocco, G., Vinson, D.P., Lewis, W. & Garrett, M.F. “Representing the meanings of objects and action words: The feature and unitary semantic space hypothesis.” *Cognitive Psychology*, 48 (2004), 422-488.

Vigliocco, G., Meteyard, L., Andrews, M. & Kousta, S. “Toward a theory of semantic representation.” *Language and Cognition*, 1 (2009), 219-240.

Vigliocco, G., Kousta, S., Vinson, D., Andrews, M. & Del Campo, E. “The representation of abstract words: What matters? Reply to Paivio's (2013) comment on Kousta et al. (2011).” *Journal of Experimental Psychology: General*, 142 (2013), 288-291.

Vigliocco, G., Vinson, D.P., Druks, J., Barber, H. & Cappa, S.F. “Nouns and verbs in the brain: A review of behavioral, electrophysiological, neuropsychological, and imaging studies.” *Neuroscience and Biobehavioral Reviews*, 35 (2011), 407-426.

Wittgenstein, Ludwig. *Tractatus Logico-Philosophicus*. 1921. Translated by D. F. Pears and B. F. McGuinness. Humanities Press. 1961.