

DOMINIK RUS
Georgetown University
Departments of Linguistics,
Psychology and Neuroscience

Washington DC, April 2004

Research training proposal:

Domain-Specific Knowledge Systems of the Brain: Evidence for the Organization of the Mental Lexicon

ABSTRACT:

Two of the most common phenomena in the acquisition, storage, and retrieval of lexical entries in the human mental lexicon have centered on the notions of *(lexical-)semantic networks* and *lexical feature decomposition*. Behavioral data from numerous case and brain lesion studies have provided massive evidence that the mental lexicon is organized *modularly* along the lines of some *semantic nets*, with specific brain regions probably corresponding to specific *semantic features* or *categories*. Besides behavioral psycholinguistic and psychological research one is also tempted to study the functional architecture of the mental lexicon from a neuroscience perspective. Such research usually stresses the importance of discriminated patterns of response to certain categories of words (or classes of words) in the human brain, trying to shed light onto the question of how *cognitive and neural mechanisms underlying category-specific knowledge are organized*, how they interact and if there are *specific neuroanatomical areas* that would store specific semantic meanings. Three working hypotheses have been offered in the cognitive neuroscience literature in the 90s and beyond on the neural basis for category-specific knowledge. The *Distributed Hypothesis* (DH), the *Sensory-Motor Hypothesis* (SM), and the *Categorical Hypothesis* (CG). Our study is an example of a *large-scale study* that includes healthy subjects as well as subjects with brain lesions, healthy subjects coming from different background with different expertise. Using several visual stimuli with different details in semantic features and of different semantic categories, we provide a mixed account between SM and DH, refuting CG, further claiming that a concept of a lexical entity, be it concrete or abstract, does not reside in a cortical 'compartment' as a single static node, but rather that the storage and retrieval of knowledge for lexical entities depends on *partially segregated large-scale neural systems*, similar to Tranel et al.'s hypothesis. However, we believe that those modules are *flexible and modifiable*, dependent on *background, learning, culture, and expertise* and are not the same across the board in all humans. We claim that a concept presented either orthographically or pictorially is recognized as an *aggregate of visual-perceptual features*, such as color or shape, as well as other *decomposition semantic features such as form or function*. Hence, the activated regions for specific categories have a sort of *mediatory function* in lexical retrieval so that when a concept of a tool is evoked (activating the sensorimotor regions that support conceptual knowledge of that tool's function and use), other, intermediary region becomes active as well, promoting the explicit phonemic knowledge pertaining to the word that denotes a given tool (in the sense of Damasio et al. 1996 or Tranel et al. 1997). We do expect stronger activation in *left temporal cortex* for animals, *left prefrontal cortex* for objects and significantly greater (presumably bilateral) activation with abstract nouns, probably due to *greater cognitive resources* used while retrieving abstract nouns. We also claim that *the number and type of features can vary* from subject to subject, though there must be a minimal set of some primitive components of a given entity that makes up that entity. We predict that the set of features may/will become smaller with expertise and familiarity though still activating similar/the same cortical regions.

I. SPECIFIC AIMS

Two of the most classic and common phenomena in the acquisition, storage, and retrieval of lexical entries in the human mental lexicon have centered on the notions of *(lexical-)semantic networks* and *lexical feature decomposition*. A lexical-semantic network refers to the network of nodes with connection strengths or distances that reflect semantic similarity among lexical items. Such networks, advocated particularly by cognitive psychologists and computational psychologists (Collins & Loftus 1975; Fellbaum 1998; Marslen-Wilson 1989, Morton 1979, Neely 1991, cited in Lowe 1997) view the human lexicon as a neural network of nodes, each represented by a word with nodes having different activation levels, and posit that when the node's strength is adequate enough, it will fire and the stimulus will be recognized as a word. For example, if a word 'strawberry' is presented shortly before 'fruit', the node associated with 'fruit' will reach threshold faster and fire sooner. Though the theory of semantic networks have been conceptually and empirically attractive (most behavioral data come from semantic priming on word recognition tasks; cf. Taft 1991), no theoretical approach to date has provided a principled account of *similarity* and *relatedness*. Semanticists and computational linguists, on the other hand, have based their lexical-semantic theories of word learning and word generation on the so-called *meaning primitives* or *semantic features*. This theory relies on the assumption that the semantic meaning of an entity can be *decomposed* into a limited set of smaller meaningful units. These semantic components represent the basic primitives of the human knowledge¹. Hence 'a female runner' could be decomposed into [+animate], [+female], [+self-propelled movement], whereas 'a ball' could be decomposed as [-animate], [-self-propelled movement]².

¹ This idea dates back at least to Gottfried W. Leibniz, whose idea was that things in nature can be either simple substances (*monads*) or complex substances (*compounds*), where the former are the indivisible atoms of nature and the latter composed bodies or aggregates. The cognitive neuroscience of vision has adopted this approach in the research of object perception, where objects are believed to be built up of primitives (= *partonomy*; see Marr & Nishihara 1978, Marr 1982, Farah 2000; see also Kanwisher 2000 for face perception along the lines of object decomposition). A further distinction of semantic components has been into *a priori* or *innate* components and the *acquired concepts* of acquisition (*posteriori* components). Psychological and linguistic accounts of language development are still pretty vague in terms of the innateness and/or evolution of such features as well as the *type* of semantic components that could be plausible candidates for innate features.

² It is pretty obvious to see the obstacles that theories positing feature decomposition come across. Sometimes there is no upper bound of the number of features needed to discriminate between two or more entities. If 'SOFA' is decomposed as {[+concrete], [-animate], [+artifact], [+to rest on]}, then how does this set of primitives differs from 'BED' with the same primitives. In other words, how plausible is positing

One of the questions that has been raised in the literature and that we are also asking here is *the status and the (neuro)psychological feasibility of feature primitives* (see Fodor 1975, De Almeida 1986, cited in Haertl 2003). In other words, we believe that experimental psychology, and particularly cognitive neuroscience, can shed light into that matter of feature (de)composition and the organization of category-based knowledge in the human brain. Hence, besides behavioral psycholinguistic and psychological research one is tempted to study the functional architecture of the mental lexicon from a neuroscience perspective. Such research usually stresses the importance of discriminated patterns of response to certain categories of words (or classes of words) in the human brain. In other words, this type of research tries to shed light onto the question of how cognitive and neural mechanisms underlying *category-specific knowledge* are organized, how they interact and if there are specific neuroanatomical areas that would store specific semantic meanings.

Our research will bridge cognitive and experimental psychology with cognitive neuroscience, addressing the question of how category-specific knowledge is stored in the human brain and how semantic feature decomposition theory can be viewed from a neurological perspective, giving rise to psychological reality. The research has both theoretical and practical implications. It can contribute to the theoretical models of the mental lexicon and its functional and anatomical organization as well as the clinical application (e.g., in speech therapy with aphasics, Alzheimer patients, or other patients suffering from category-specific semantic memory loss)³. Furthermore, our research can provide some insights and speculations about evolutionary pressures associated with survival of the species that has developed a dedicated neural mechanism for processing specific kinds of knowledge.

II. BACKGROUND AND SIGNIFICANCE

Though brain imaging techniques are pretty recent additions to the study of neural correlates of category-specific knowledge, cognitive and experimental psychology in the 70s already offered

[+has arms] as a feature to distinguish a sofa from a bed. Also, would [+kind] need to be a semantic primitive to differentiate between 'WOMAN' and 'WITCH' or 'MOTHER-IN-LAW' (presumably all with {[+animate], [+human], [+female]}).

³ Category-specific disorders normally result from brain damage following stroke, head injury, herpes simplex, viral encephalitis or dementia of the Alzheimer type.

intriguing behavioral data suggesting that patients with certain brain damages show selective loss of knowledge about a specific category of objects. The pioneering study was that of Goodglass and colleagues (Goodglass et al. 1976), which examined the production and comprehension of various patients for several categories, including *numbers*, *letters*, *actions*, *colors*, and *objects*. Ever since then the research into the functional semantic segregation within the human brain has flourished considerably. A number of detailed case studies have found a number of selective semantic deficits for concepts as specific as *fruits* and *vegetables* (Hart, Berndt & Caramazza 1991), *body parts* (Dennis 1976), *colors* (Geshwin & Fusillo 1966; cited in Martin et al. 2000), *geographical names* (McKenna & Warrington 1976), and *proper names* (Semensa & Zettin 1988, 1989; cited in Gazzaniga 2000). Though it is sometimes hard to accurately interpret this behavioral data (most of which is based on RT tasks on item recognition), all these studies suggest that the organization of semantic concepts in the brain is *highly modular*. Most of the imaging research of the 90s, however, has concentrated on the distinction in the mental representation of *living (human) vs. non-living (non-human)* entities, confirming the classic behavioral studies of Warrington & Shallice (1984) and Silveri & Gainotti (1988), which described patients who were unable to identify or name *people* and other *living things*, and of Warrington & McCarthy (1987), who described patients who were unable to recognize *man-made objects*⁴. Such selective loss of semantic knowledge was intriguing and puzzling in those days, particularly due to the lack of brain imaging technology and possibly also due to restricted and less-understood models of the mental lexicon.

Three explanatory hypotheses have been offered in the cognitive neuroscience literature in the 90s and beyond on the neural basis for category-specific knowledge. *The Distributed Hypothesis* (DH) views categorical knowledge as distributed throughout the cerebral cortex, positing that category-specific effects reflect the differences due to the extent to which semantic primitives are shared and intercorrelated. This view does not posit segregated semantic representations, but proposes that

⁴ The most common categorical distinction in the imaging studies from the 90s include animals vs. tools (e.g., Martin et al. 1996); people vs. animals vs. tools (e.g., Damasio et al. 1996); animals & fruits vs. vehicles & tools (e.g., Moore & Price 1999); living things (people, animals, plants) vs. artifacts (materials, tools) (e.g., Mummery et al. 1998); animals & fruits & vegetables vs. tools & utensils (Tranel et al. 1997); for a review of more studies see Grossman et al. 2002

different types of objects place different demands on a shared processing system. Hence, preserved semantic knowledge within a category with many primitive features can compensate for some knowledge loss. For instance, in the case of animals, the frequent co-occurrence of *a tail* and *four legs* can compensate for degraded knowledge about a specific category member, such as loss of knowledge that the concept of a *cow* has a tail. In contrast, objects such as tools share fewer intercorrelated features and the loss of knowledge that a *saw* has a serrated blade cannot be compensated for by knowing that it has a handle or by intact knowledge about other tools (Tyler et al. 1996, 2000; see Grossman et al. 2002 for a review). *The Sensory-Motor Hypothesis* (SMH) posits neuroanatomical localization of modality-specific feature knowledge. Under this view, cortical regions in which semantic feature knowledge is stored reflect its most relevant modality (either visual or motor). Hence, visual-perceptual information is believed to contribute disproportionately to the features associated with natural kinds and the breakdown in the visual-perceptual cortices in left temporal and occipital regions has compromised comprehension and naming of natural kinds relative to manufactured objects (Damasio et al. 1996, Hart et al. 1998, Tranel et al. 1997). *The Categorical Hypothesis* (CH) posits that object knowledge is ‘compartmentalized’ and localized brain damage can cause selective impairments within a modular system (Caramazza & Shelton 1998).

None of the imaging studies up to date survives detailed scrutiny. When collapsed together, the imaging studies seem to be very contradicting, having posited activations in different brain regions in both left and right hemispheres in temporal, occipital and parietal lobes (cf. Table 1 in Moore & Price 1999). The first problem that arises in these studies is the distinction between *recognition* and *naming*, which are cognitively and neurally distinct processes, though most researchers use them inconsistently or even interchangeably. Second, most studies have used only familiar objects and have not controlled for familiarity with the objects tested nor for the subjects who would be more or less familiar with certain objects. Third, the materials used in most studies were black line drawings. This abstracts away a lot of primitive features of an entity, such as [+furry] for certain animal exemplars, for example (considering these are primitive features that compose entities in the first place, of

course). Furthermore, the number of categories used in imaging studies has been extremely small, focusing mainly on animals, tools and people. Furthermore, almost all entities have been concrete nouns in the absence of abstract nouns (or verbs or even other word classes). Last but not least, conceptually we would not want to say that a concept for a certain entity ‘resides’ where its name is stored. Understanding the meaning of the word does not occur in a processing vacuum, but requires *meaning-related processes* that link a word to its network of associates and to additional information crucial to comprehension⁵.

To control for all these variables, we propose a *large-scale study* that will include healthy subjects as well as subjects with brain lesions (to control for anatomical accuracy), high school students with no particular ‘expertise’ in tools, animals or fruits/vegetables, as well as 20 Home Depot employees with extensive knowledge of and familiarity with tools, 20 veterinarians with familiarity with animals and 20 greengrocers with familiarity of fruits and vegetables (to control for familiarity). We will test all these subjects with all three procedures that have appeared in the literature (a) recognition of pictures/typed words, b) naming of pictures, and c) picture matching (to control for processing difficulty and modality-related processing), with black line drawings, colored drawings as well as multi-feature drawings (photographs) (to control for feature composition & the claims of the DH), and a large(r) number of categories (animals, fruits, vegetables/unique & non-unique people/tools and utensils, furniture; as well as abstract nouns, to control for the abstract-non-abstract distinction as well as possible differences within a single category, e.g., ‘living things’).

We believe that a concept of a lexical entity, be it concrete or abstract, does not reside in a cortical ‘compartment’ as a single static node (refuting the CH), but rather that the storage and retrieval of knowledge for lexical entities depends on partially segregated large-scale neural systems, similar to Tranel et al.’s hypothesis. However, we believe that those modules are flexible and modifiable, dependent on background, learning, culture and expertise and are not the same across the board in all

⁵ This is how the mental lexicon is traditionally viewed linguistically and computationally (cf. Lowe 1997, Gazzaniga 2000 among others). The mental lexicon refers to the representations that allow word recognition on the basis of auditory and visual stimuli. The lexicon is hence understood as linking two levels of representation, a form-based representation that contains phonological (or graphemic properties) and a semantic representation that reflects the meaning of a particular lexical item in relation to other items.

humans (hence, we will also look at individual differences). We claim that a concept presented either orthographically or pictorially is recognized as an aggregate of visual-perceptual features, such as color or shape, as well as other decomposition semantic features such as form or function. Hence, the activated regions for specific categories have a sort of mediatory function in lexical retrieval so that when a concept of a tool is evoked (activating the sensorimotor regions that support conceptual knowledge of that tool's function and use), other, intermediary region becomes active as well, promoting the explicit phonemic knowledge pertaining to the word that denotes a given tool (in the sense of Damasio et al. 1996 or Tranel et al. 1997). We do expect stronger activation in left temporal cortex for animals, left prefrontal cortex for objects and significantly greater (presumably bilateral) activation with abstract nouns, probably due to greater cognitive resources (Grossman et al. 2002). We also claim that the number and type of features can vary from a subject to subject though there must be a minimal set of some primitive components of a given entity that makes up that entity. This will be tested through material manipulation (black lines vs. color vs. real-life photographs). We predict that the set of features may/will become smaller with expertise and familiarity though still activating similar/the same cortical regions.

III. EXPERIMENTAL DESIGN AND METHODS

Predictions: see above

Experiments: 1) behavioral (RTs): picture recognition, picture naming, word recognition, picture matching
2) fMRI w/ same tasks
3) PET w/ same tasks

Participants: right-handed English native speakers; 20 patients w/ brain lesions, 20 Home Depot employees, 20 veterinarians, 20 greengrocers, 20 high school students (one group of healthy subjects serves as control for the other group)

Materials: 1) pictures of categories outlined above; variables: black lines vs. color vs. real-life photographs; orthographic words matched for mean-length and frequency (Francis & Kucera 1982); see above for categories testes

- Clinical applications:** in speech therapy, part. for patients with Alzheimer's disease, patients with brain injuries and brain damage who have problems with semantic memory, associated w/ specific semantic categories
- Length of study:** 1-2 years

IV. LITERATURE CITATIONS

- Caramazza, A. et al. 1990. The multiple semantics hypothesis: Multiple confusions? *Cognitive Neuropsychology* 7: 161-189.
- Caramazza, A. & Shelton, J. 1998. Domain-Specific Knowledge Systems in the Brain: The Animate-Inanimate Distinction. *Journal of Cognitive Neuroscience* 10: 1-34.
- Collins, A.M. & Loftus, E.F. A spreading activation theory of semantic processing. *Psychological Review* 82: 407-428.
- Damasio, H. et al. 1996. A neural basis for lexical retrieval. *Nature* 380: 499-505.
- Dennis, M. 1976. Dissociated naming and locating of body parts after left anterior temporal lobe resection: An experimental case study. *Brain and Language* 3: 147-163.
- Goodglass H. et al. The "tip-of-the-tongue" phenomenon in aphasia. *Cortex*, 12: 145-153, 1976.
- Fellbaum, C. 1998. *Wordnet: An Electronic Lexical Database*. Cambridge, MA: The MIT Press.
- Gazzaniga, M.S. Ed. *The New Cognitive Neurosciences* (2nd edition). MIT Press, 2000
- Grossman, M. et al. 2002. The neural Basis for Category-Specific Knowledge: An fMRI Study. *Neuroimage* 15: 936-848.
- Hart, J. Berndt, R. & Caramazza, A. 1985. Category-specific naming deficit following cerebral infarction. *Nature* 316: 439-440.
- Haxby et al. 2001. Distributed and Overlapping representations of Faces and Objects in Ventral Temporal Cortex. *SCIENCE* 293: 2425-2430.
- Haertl, D. 2003. *The Mental Lexicon*. Class Notes. Retrieved at http://www2.hu-berlin.de/engl/haertl/haertl_lexicon.html
- Levelt, W.J.M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22: 1-38.
- Lowe, W. 1997. *Meaning and the Mental Lexicon*. Ms. Centre for Cognitive Science. University of Edinburgh.
- Martin et al. 1996. Neural correlates of category-specific knowledge. *Nature*, 379: 649-652.
- McCarthy, R.A., & Warrington, E.K. Evidence for modality-specific meaning systems in the brain. *Nature* 334: 428-430, 1988.
- Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90: 227-234.
- Moore, C.J. & Price, C.J. 1999. A functional neuroimaging study of the variables that generate category-specific object processing differences. *Brain* 122: 943-962.
- Mummery et al. 1998. Functional neuroanatomy of the semantic system: divisible by what? *Journal of Cognitive Neuroscience* 10: 766-777.
- Silveri, M.C. & Gainotti, G. 1988. Interaction between vision and language in category-specific semantic impairment. *Cognitive Neuropsychology* 5: 677-709.
- Taft, M. 1991. *Reading and the Mental Lexicon*. Hove: LEA
- Tranel, D., Damasio H. & Damasio A.R. 1997. A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia* Vol. 35, No. 10: 319-1327.
- Warrington, E.K., and McCarthy, R. 1987. Categories of knowledge: Further fractionations and an attempted integration. *Brain* 110: 1273-1296.
- Warrington, E.K., and Shallice, T. 1984. Category specific semantic impairments. *Brain* 107: 829-254.