On Discriminating fMRI Representations of Abstract WordNet Taxonomic Categories

Andrew James ANDERSON[1]*,Yuan TAO*[1]*, Brian MURPHY*[2]*, Massimo POESIO*[1,3]
(1) Centro Interdipartimentale Mente e Cervello (CIMeC), University of Trento, Italy
(2) Machine Learning Department, School of Computer Science, Carnegie Mellon University, USA
(3) School of Computer Science and Electronic Engineering, University of Essex, UK
andrew.anderson@unitn.it, yuan.tao@unitn.it, brianmurphy@cmu.edu, massimo.poesio@unitn.it

ABSTRACT

How abstract knowledge is organised is a key question in cognitive science, and has clear repercussions for the design of artifical lexical resources, but is poorly understood. We present fMRI results for an experiment where participants imagined situations associated with abstract words, when cued with a visual word stimulus. We use a multivariate-pattern analysis procedure to demonstrate that 7 WordNet style Taxonomic categories (e.g. 'Attribute', 'Event', 'Social-Role'), can be decoded from neural data at a level better than chance. This demonstrates that category distinctions in artificial lexical resources have some explanatory value for neural organisation.

Secondly, we tested for similarity in the interrelationship of the taxonomic categories in our fMRI data and the associated interrelations in popular distributed semantic models (LSA,HAL,COALS). Although distributed models have been successfully applied to predict concrete noun fMRI data (e.g. Mitchell et al., 2008), no evidence of association was found for our abstract concepts. This suggests that development of new models/experimental strategies may be necessary to elucidate the organisation of abstract knowledge.

KEYWORDS : fMRI, CONCEPT REPRESENTATION, ABSTRACT, MVPA, WORDNET

1 Introduction

Data about the organization of conceptual knowledge in the brain coming from patients with semantic deficits (e.g. Warrington & Shallice, 1984, Caramazza & Shelton, 1998) or collected from healthy patients using functional Magnetic Resonance Imaging[1] (fMRI) (e.g. Martin & Chao, 2001) have proven an essential source of evidence for our understanding of conceptual representations, particularly when analyzed using machine learning methods (e.g. Haxby et al 2001, Mitchell et al., 2008). Most of this work has focused on a fairly narrow range of conceptual categories, primarily concrete concepts such as animals, plants, tools, etc., which represent only a small percentage of the range of conceptual categories that are part of human knowledge. Until recently only a few studies studied the representation in the brain of abstract concepts such as law

[1] functional Magnetic Resonance Imaging measures blood flow in the brain, which reflects neural cells' energy consumption which in turn is generally regarded to relate to neural activity. Comparative to other popular neuroimaging techniques (e.g. EEG, MEG) fMRI offers relatively high spatial resolution (data is measured as a 3D volume built from rectangular cuboids known as voxels, of side 1-5 mm, over the entire brain) at relatively low sampling frequency (commonly ≥ 1Hz).

or freedom (Binder et al, 2005; Friederici et al, 2002; Grossman et al, 2002). Some recent studies have shown that fMRI data contain sufficient information to discriminate between concrete and abstract concepts (Binder et al, 2005; Wang et al, 2012) but meta-analyses such as (Wang et al, 2010) also showed that fairly different results are obtained depending on the types of abstract concepts under study, and that the range of abstract concepts considered tends to be fairly narrow.

This type of analysis is complicated by the fact that the representation and organization of human knowledge about abstract conceptual categories is much less understood than for concrete concepts. Human intuitions about abstract concepts are not very sharp: e.g., studies asking subjects to specify the defining characteristics of abstract concepts find that this task is much harder than for concrete ones (Hampton 1981, McRae & Cree, 2002, Wiemer-Hastings & Xu, 2005). On the theoretical side, as well, there is not much agreement on abstract concepts among psychologists, (computational) linguists, philosophers and other cognitive scientists who have proposed theories about the organization of conceptual knowledge. Just about the only point of agreement among such proposals is that there is no such thing as an 'abstract concept' –human conceptual knowledge includes a great variety of abstract categories of varying degrees of abstractness ranging from knowledge about space and time (e.g., day, country) to knowledge about actions and events (e.g., solo, robbery) to knowledge about inner states including emotions (fear) and cognitive states (belief), to purely abstract concepts (e.g., art, jazz, law). It is also known that many of these categories have their own distinct representation in memory (Binder & Desai, 2009). But there is a lot of disagreement among exactly which categories these different types of abstract concepts belong to, e.g., which category does the concept law belong to. These disagreements are clearly in evidence in the significant differences between the representation of such categories in the large-scale repositories of conceptual knowledge that have been developed in the last twenty years, such as WordNet (Fellbaum, 1998), CYC (Lenat, & Guha, 1990) and DOLCE (Gangemi et al, 2002). In WordNet, the top category 'abstract concept' covers attributes, events and actions, temporal entities, and highly abstract concepts such as law both in the sense of 'collection of all laws' and in the sense of 'area of study', whereas locations are considered concrete concepts. In DOLCE, actions and events, attributes, and highly abstract concepts such as propositions are treated as completely unrelated conceptual categories, whereas both temporal and spatial locations are included in the quality category.

It follows that there is joint motivation from cognitive science and computational linguistics to extend our understanding of abstract knowledge representation. The objectives of the present work are two fold, (1) to broaden the range of abstract concepts studied using neuroimaging; (2) to examine whether artificial knowledge representation strategies can be used to interpret fMRI data.

We adopt an fMRI paradigm, where stimuli were presented in the form of words on the screen and participants were required to imagine a situation associated with the word. We used as stimuli concepts belonging to seven distinct WordNet style taxonomic categories, ranging from concrete to more abstract (tool, location, social role, event, communication, attribute, and a category we called urabstract of highly abstract words) and two different domains (music and law). Domain membership is not important to this paper and will be addressed in future work (this point is returned to in section 4). Firstly a Multivariate Pattern Analysis (MVPA) procedure was used to test whether single stimulus trials could be classified by their taxonomic class. On demonstrating that classifications can indeed be made at a level better than chance (section 3.1), we further examined whether there are similarities between concept representations in the fMRI data and popular distributed semantic models used in computational linguistics (section 3.2).

Three semantic models were selected: Hyperspace Analogue to Language (HAL) (Burgess, 1998), Correlated Occurrence Analogue to Lexical Semantics (COALS) (Rohde, et al., 2005) which is a refinement of HAL and Latent Semantic Analysis (LSA) (Landauer et al, 1998). All three models express meaning in terms of a multidimensional statistical model of a word's context. HAL models meaning as a function of the number of times a word occurs in close proximity to a each of a large set of feature words, within a large body of text. LSA counts the occurrences of words in individual documents and subsequently reduces the dimensionality (in documents) through singular value decomposition. COALS incorporates a number of algorithmic modifications to the HAL, including data reduction by singular value decomposition. The important conceptual difference is that LSA attempts to bind words to topic (assumed to be derived from the general themes of the documents), whereas HAL and COALS captures meaning through word inter-relations. All models have been applied with success in one way or other to interpret human cognition in a variety of semantic tasks and psychological experiments, including synonym test, word relatedness judgment, semantic priming, semantic categorization, (Lund & Burgess, 1996; Burgess, 1998; Landauer et al., 1997, 1998; Rohde et al., 2005). Despite their success in explaining behavioural tasks, by using representational dissimilarity analysis (section 3.3) we found that none of the models provide a good general match for the structure of the abstract fMRI data.

2 Methods

2.1 Participants

Seven right handed native Italian speakers (3 female), aged between 19 and 38, were recruited to take part in the study. All had normal or corrected-to-normal vision. Participants received compensation of €15 per hour. The studies were conducted under the approval of the ethics committee of the host University, and participants gave informed consent.

2.2 Data Acquisition

fMRI images were recorded on a 4T Bruker MedSpec MRI. An EPI pulse sequence with TR=1000ms, TE=33ms, and 26° flip angle was used. A 64 * 64 acquisition matrix was used and seventeen slices were imaged with a between slice gap of 1mm. Voxels had dimensions 3mm * 3mm * 5mm.

2.3 Experimental Paradigm

The names of 70 concepts were presented to participants in the form of written words on the screen. The stimuli were displayed using bold Arial-Black size 20 font on a grey background. Each stimulus was presented five times, for a total of 350 trials, split in five blocks with the order of presentation being randomized in each block. Participants had the opportunity to pause between blocks and the overall task time did not exceed 60 minutes. Each trial began with the presentation of a blank screen for 0.5s, followed by the stimulus word of dark grey on a light grey background for 3s, and a fixation cross for 6.5s. Participants were asked to keep still during the task and during breaks.

With concrete concepts, participants are often asked to think actively about the properties of the object named (see, e.g., Mitchell et al, 2008) but eliciting properties is not so easy for abstract concepts. On the other hand, participants to studies such as (Hampton, 1981; McRae & Cree, 2002; Wiemer-Hastings & Xu, 2005) appeared able to produce situation-related objects. Our participants were therefore instructed to "think about situations that exemplify the object the word refers to".

The list of concept words were supplied to participants in advance of the experiment, so that they could prepare appropriate situations to simulate consistently.

2.4 Materials

Our objective was to obtain a list of words representative of the full range of non-concrete concepts. The list of categories was produced by associating WordNet (Fellbaum, 1998) categories to the terms with highest abstractness ranking in an abstractness norm for Italian. We identified the 6 WordNet categories that occurred most frequently in the norms. Finally, WordNet Domains (Pianta et al, 2002) was used to select 70 words whose unique or most preferred sense belonged to these categories.

More in detail, our starting point was the set of behavioural norms by Barca et al (2002) listing Italian words ranked by perceived abstractness. These words were next looked up in the Italian WordNet contained in MultiWordNet (Pianta et al, 2002) to determine the taxonomic category of their dominant sense(s). The authors edited this list down to a set of six taxonomic categories of concepts found in Barca et al's norms plus a category of concrete concepts, *tool*, for comparison purposes. The six non-concrete categories are:

Locations, including concepts such as court, jail and theatre. *Locations* are considered as concrete objects in WordNet but belong to the separate category `qualities' in DOLCE, and could therefore be considered concepts in between concrete and abstract.

Four non-concrete categories of arguably increasing levels of abstractness: *event*, *communication* (covering concepts such as accusation or symphony), *attribute*, and *urabstract* (our term for concepts such as law or jazz which are fairly common in abstractness norms, are classified as abstract in WordNet, but do not belong to a clear subcategory of abstract such as event or attribute)

Finally, the category *social-role*, containing concepts such as judge or tenor which are fairly common in abstractness norms and are typically associated with scenarios but whose status as concrete or abstract is not very clear. The complete word list including English translations of the Italian stimuli is in TABLE 1.

2.5 Preprocessing

Preprocessing was undertaken using the Statistical Parametric Mapping software (SPM99, Wellcome Department of Cognitive Neurology, London, UK). Data were corrected for head motion, unwarped (to compensate for geometric distortions in the image interacting with motion) and spatially normalised to the MNI template image and resampled at 3mm * 3mm * 6mm. Only voxels estimated to be grey matter were included in the subsequent analysis. For each participant the data, per voxel, in each session (presentation cycle of 70 words) was corrected for linear trend and transformed to z-scores.

A single volume was computed to represent each stimulus word, by taking the voxel-wise mean of the four seconds of data offset by four seconds from the stimulus onset (to account for hemodynamic response).

tool	manette	handcuffs	violino	violin
	toga	robe	tamburo	drum
	manganello	truncheon	tromba	trumpet
	cappio	noose	metronomo	metronome
	grimaldello	skeleton key	radio	radio
location	tribunale	court/tribunal	palco	stage
	carcere	prison	auditorium	auditorium
	questura	police station	discoteca	disco
	penitenziario	penitentiary	conservatorio	conservatory
	patibolo	gallows	teatro	theatre
social-role	giudice	judge	musicista	musician
	ladro	thief	cantante	singer
	imputato	defendant	compositore	composer
	testimone	witness	chitarrista	guitarist
	avvocato	lawyer	tenore	tenor
event	arresto	arrest	concerto	concert
	processo	trial	recital	recital
	reato	crime	assolo	solo
	furto	theft	festival	festival
	assoluzione	acquittal	spettacolo	show
communication	divieto	prohibition	canzone	song
	verdetto	verdict	pentagramma	stave
	ordinanza	decree	ballata	ballad
	addebito	accusation	ritornello	refrain
	ingiunzione	injunction	sinfonia	symphony
attribute	giurisdizione	jurisdiction	sonorita'	sonority
	cittadinanza	citizenship	ritmo	rhythm
	impunita'	impunity	melodia	melody
	legalita'	legality	tonalita'	tonality
	illegalita'	illegality	intonazione	pitch
urabstracts	giustizia	justice	musica	music
	liberta'	liberty	blues	blues
	legge	law	jazz	jazz
	corruzione	corruption	canto	singing
	refurtiva	loot	punk	punk

TABLE 1. Italian stimuli words and English translations, Taxonomic category is indicated in the left column. Taxonomic categories are ordered in terms of increasing abstractness.

2.6 Cross validation analysis procedure

Broadly the same cross-validation procedure was followed for each analyses. Input and target data pairs were partitioned into training and testing sets (using a leave-n-out approach) to support a number of cross validation iterations. Target patterns were binary vectors with a single field set to one to uniquely specify the category. Input was a masked version of the fMRI grey-matter data, retaining the 1000 most stable voxels in the training set according to the following procedure, similar to that used by Mitchell et al. (2008). For each voxel, the set of 70 words from each unique pair of scanning sessions in the training set were correlated, and the mean of the six resulting correlations (from 4 scanning sets) was taken as the measure of stability. The 1000 voxels with highest mean correlations were selected for analysis.

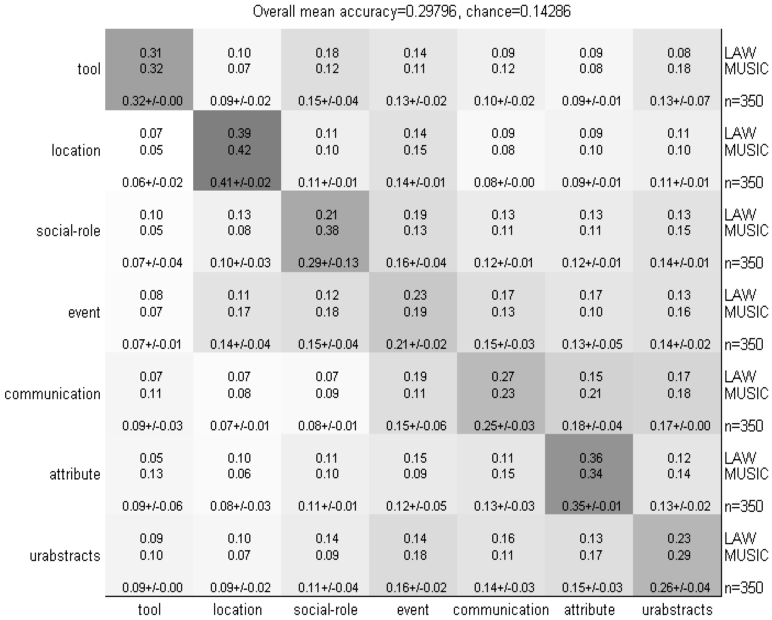
Pattern classification used a single layer neural network with logistic activation functions (MATLAB 2009B, Mathworks, Neural Network toolbox). Weights and biases were initialized using the Nguyen-Widrow algorithm and training used conjugate gradient decent, continued until convergence, with performance evaluated using mean square error, with a goal of 10^{-4} or completion of 2000 training epochs. In each cross-validation iteration the network was trained using the masked fMRI data and binary target codes in the training set and subsequently tested on the previously unseen masked fMRI data. The Euclidean distance between the network output vectors and target codes was computed, and the target code with the minimum distance selected as the network output.

3 Results

Leave-out-session cross validation analyses were undertaken for each participant to recognize taxonomic distinctions from the fMRI data. There were 5 scanning sessions, therefore training in each of the five cross-validation iterations was on 280 words (4 replicates of each of the 70 stimulus words) and testing was on the remaining 70 words. Figure 1 shows a confusion matrix averaging results across all 7 participants (and cross-validation iterations within participant).

3.1 Can taxonomic distinctions be recognized within participant?

Mean classification accuracy for the 7-way taxonomic distinctions was ~0.3 with chance level at 0.143. Accuracy is greatest for location, tool and attributes and there is a visible diagonal in Figure 1, suggesting all classes can be discriminated. This claim is however statistically unsubstantiated, and indeed until recently the question of how to rigorously interpret the classification performance of multiway classifiers had not been directly addressed. Binomial tests are often applied to test whether a classifier is predicting randomly, however in the multi-class case this leaves many questions unanswered. For instance, here there were 730/2450 correct classifications, and the probability of achieving this by chance is $p=2.2*10^{-16}$ (2-tailed Binomial test), however this does not answer whether the classifier capable of distinguishing between all test categories, or just between subsets of categories. Motivated by these concerns, and drawing from the statistical literature of contingency tables, Olivetti et al (2012) developed a test exploiting Bayesian hypothesis testing to evaluate the posterior probability of each possible partitioning of distinguishable subsets of test classes. For example taking three classes, possible distinguishable test class partitions are [1][2][3]; [1,2][3]; [1,3][2]; [1][2,3]; [1,2,3], and each of these would be assigned a posterior probability, where as a general rule of thumb a probability in excess of 1/K, where K is the number of hypotheses, (i.e., 5 in the 3 class example) would be seen as informative evidence. (Olivetti pers. comm.)



Overall mean accuracy=0.29796, chance=0.14286

	tool	location	social-role	event	communication	attribute	urabstracts	
tool	0.31 0.32 0.32+/-0.00	0.10 0.07 0.09+/-0.02	0.18 0.12 0.15+/-0.04	0.14 0.11 0.13+/-0.02	0.09 0.12 0.10+/-0.02	0.09 0.08 0.09+/-0.01	0.08 0.18 0.13+/-0.07	LAW MUSIC n=350
location	0.07 0.05 0.06+/-0.02	0.39 0.42 0.41+/-0.02	0.11 0.10 0.11+/-0.01	0.14 0.15 0.14+/-0.01	0.09 0.08 0.08+/-0.00	0.09 0.10 0.09+/-0.01	0.11 0.10 0.11+/-0.01	LAW MUSIC n=350
social-role	0.10 0.05 0.07+/-0.04	0.13 0.08 0.10+/-0.03	0.21 0.38 0.29+/-0.13	0.19 0.13 0.16+/-0.04	0.13 0.11 0.12+/-0.01	0.13 0.11 0.12+/-0.01	0.13 0.15 0.14+/-0.01	LAW MUSIC n=350
event	0.08 0.07 0.07+/-0.01	0.11 0.17 0.14+/-0.04	0.12 0.18 0.15+/-0.04	0.23 0.19 0.21+/-0.02	0.17 0.13 0.15+/-0.03	0.17 0.10 0.13+/-0.05	0.13 0.16 0.14+/-0.02	LAW MUSIC n=350
communication	0.07 0.11 0.09+/-0.03	0.07 0.08 0.07+/-0.01	0.07 0.09 0.08+/-0.01	0.19 0.11 0.15+/-0.06	0.27 0.23 0.25+/-0.03	0.15 0.21 0.18+/-0.04	0.17 0.18 0.17+/-0.00	LAW MUSIC n=350
attribute	0.05 0.13 0.09+/-0.06	0.10 0.06 0.08+/-0.03	0.11 0.10 0.11+/-0.01	0.15 0.09 0.12+/-0.05	0.11 0.15 0.13+/-0.03	0.36 0.34 0.35+/-0.01	0.12 0.14 0.13+/-0.02	LAW MUSIC n=350
urabstracts	0.09 0.10 0.09+/-0.00	0.10 0.07 0.09+/-0.02	0.14 0.09 0.11+/-0.04	0.14 0.18 0.16+/-0.02	0.16 0.11 0.14+/-0.03	0.13 0.17 0.15+/-0.03	0.23 0.29 0.26+/-0.04	LAW MUSIC n=350

FIGURE 1. Leave-out-one-session Taxonomic category classification confusion matrix. Rows are the target labels and columns are predictions. Numbers overlaid on each cell indicate the proportion of predictions per law and music respectively (as indicated on the right y-axis) for that row, averaging over 7 participants. The numbers on the bottom line of each cell are the mean and standard deviation of predictions. Cell shading is scaled to the range 0 to 0.41 (0.41 is the maximum mean accuracy per cell displayed).

Applying Olivetti et al.s' (2012) test to the taxonomic confusion matrix in Figure 1 and sorting all subset partitions in descending order of posterior probability, finds the top ranking partition (posterior probability=0.93) to be that all test classes are discriminable. The highest ranked three partitions are below (posterior probabilities rapidly diminish in the remaining 874 partitions that are not displayed).

[1=tool][2=location][3=social-role][4=event][5=communication][6=attribute][7=urabstracts]

Partition: [[1][2][3][4][5][6][7]], postP: 0.93

Partition: [[1][2][3][4 5][6][7]], postP: 0.04

Partition: [[1][2][3][4 7][5][6]], postP: 0.02

Tool, Location and Attribute are most clearly distinguished, whereas prediction of taxonomic category is weakest for categories toward the middle of the concreteness scale (Event and Communication) and in the second partition of Olivetti et al.s' (2012) analysis these categories aggregate (although the posterior probability for this partition at 0.04 is much lower than the first).

3.2 Representational dissimilarity analysis between fMRI data and distributed semantic models

Representational dissimilarity analyses (Kriegeskorte, 2008) between the fMRI data and the three distributed semantic models (LSA, HAL, COALS) identified in the introduction were run to test for association in inter-representations of taxonomic classes between modalities. Each semantic model was built using the corpus itWaC. This corpus is from WaCky, a collection of very large (>1 billion words) corpora built by web crawling, and annotated with Part-of-Speech tagging and lemmatisation. itWaC is the largest publicly documented Italian language resource (Baroni *et al.*, 2008).

Representational dissimilarity analysis was as follows. For each participant, all fMRI representations within each of the seven taxonomic categories were voxel-wise averaged. Then the pairwise difference between each unique taxonomic category pairing was computed (n=21) using 1-rho as a distance metric, where rho is Spearman's rank correlation coefficient. Likewise, for LSA, HAL and COALS, semantic representations of all word models within each taxonomic category were averaged, and pairwise differences between all unique category pairs taken. The list of respective category pair differences for imaging data and each of the semantic models were correlated using Spearman's rank correlation to give a correlation coefficient for each. Following this the 7 per participant lists of 21 category pair differences were collapsed (by averaging) and the resulting list of average differences correlated with the 3 semantic models. Significance was tested using a permutation test as follows. The seven taxonomic condition labels were shuffled in every possible way to construct a null distribution that the two dissimilarity lists are not correlated. The p-value is calculated as the proportion of random correlation coefficients that are greater than or equal to the observed coefficient. Results are in Table 2.

Although there are two participants who show signs of a correlation with the HAL, HAL/COALS models, it is clear that this is not a general pattern across participants. Correlations range from positive to negative, and if p-values are corrected for multiple comparisons using Bonferroni correction (where the conventional significance threshold becomes p=0.05/21), results that individually are significant disappear. There is additionally no correlation between the fMRI dissimilarity matrices averaged over participants and the three semantic models.

4 Discussion

We have collected evidence that fMRI recordings contain sufficient information to discriminate between all Taxonomic categories that we tested. In other words, the distinctions between types of non-concrete concepts proposed in state-of-the-art models of conceptual knowledge such as WordNet are supported to a certain extent by brain data.

Participant		HAL	COALS	LSA
19730713	**rho**	**0.3571**	**0.1416**	**-0.1649**
	P-value	0.0206	0.2061	0.7502
19820508	**rho**	**0.0662**	**-0.0896**	**0.0156**
	P-value	0.346	0.6987	0.4465
19830625	**rho**	**0.5455**	**0.5312**	**-0.1091**
	P-value	0.0347	0.0407	0.6909
19850913	**rho**	**0.0364**	**-0.1169**	**0.2169**
	P-value	0.4083	0.7744	0.17
19861211	**rho**	**-0.2494**	**-0.2649**	**-0.1805**
	P-value	0.9288	0.9683	0.7756
19891011	**rho**	**-0.2338**	**-0.0805**	**-0.039**
	P-value	0.8931	0.6568	0.5299
19920102	**rho**	**0.1273**	**0.1051**	**0.0156**
	P-value	0.2581	0.2767	0.4469
Collapsed dissimilarity matrix correlation	**rho**	**0.2455**	**0.1481**	**-0.013**
	P-value	0.1351	0.2437	0.5281

TABLE 2. Representational dissimilarity analysis between neural data and semantic models.

Whereas a number of studies have demonstrated a connection between distributional semantic models and neuroimaging data for concrete concepts (e.g. Mitchell et al, 2008; Murphy et al. 2009; Murphy et al., 2011; Chang et al., 2011), representational similarity analysis failed to find a systematic association between the inter-relationship of categories in the fMRI data and the inter-relationship of categories in distributional semantic models. There could be a number of reasons for this. Firstly, it may be that the neural organisation of abstract knowledge is in fact entirely different to the distributed semantic representations in common usage. Given that the semantic models show some explanatory power for human behavioural data, it would be unwise to discount them too quickly. Alternatively it could be that the experimental/fMRI protocol used is unfit for the challenge. As concerns the experimental protocol, abstract concepts generally speaking are more difficult to imagine than concrete objects, and the richness of the neural representations invoked in our experiment may consequently be comparatively weak. Additionally we have no guarantee that participants were compliant with the task (the only gauge on this being the ability to detect systematic patterns in a participants data). It will be valuable to consider modifying the task and if/where possible, to develop tasks that require mental manipulation of the concept in a more realistic context, where the performance of the participant can be evaluated. As concerns fMRI, it is possible that abstract concepts may be represented on a smaller spatial scale than concrete concepts, especially if they are not grounded in sensorimotor mechanisms and associated neural maps (as frequently thought to be the case for concrete concepts). Thus our whole brain analysis using large voxels may overlook pertinent features. However given the success of taxonomic category classification with the current fMRI setup, it should not be dismissed to quickly either.

This paper has thus far not directly addressed an important competing theory of concept organisation. Gentner (1981), Hampton (1981), and others found that unlike concrete concepts, abstract concepts are mostly characterized in terms of relations to other entities present in a situation. Wiemer-Hastings & Xu (2005) provided further support for this finding and proposed that abstract concepts are "anchored in situations" (Wiemer-Hastings & Xu 2005, p. 731); in a similar fashion, Barsalou (1999) argued that the representation of abstract concepts is 'framed by abstract event sequences'. This suggests a scenario-based organization for non-concrete concepts. In this type of organization, non-concrete concepts are defined in terms of their role with respect to a scenario: e.g., *law* is defined with respect to the *court* scenario, whereas *jazz* is defined with relation to a *music* scenario. In fact our experimental data set was carefully selected to allow us to begin to target this question (50% of our words are associated with Law and 50% with Music). Our preliminary analyses suggest that law and music scenarios can also be successfully decoded from the neural data. Complete results will be presented in future work.

Conclusion

Conclusions are: (1) WordNet style taxonomic categories for abstract concepts, are at least cognitively relevant in that they can be distinguished from neural data; (2) In contrast to previous findings for concrete concepts, we were unable to detect a relationship between inter-representation of abstract concept categories in fMRI data and inter-representations in popular distributed semantic models.

The question of how abstract knowledge organised remains murky, however given the taxonomic classification success we are optimistic that advances are possible with current technology and methods.

References

Barca, L., Burani, C., Arduino, S., (2002). Word naming times and psycholinguistic norms for Italian nouns. *Behavior Research Methods*, 34(3): 424-434.

Baroni, M., Bernardini, S., Ferraresi, A., Zanchetta, E., (2008). The WaCky Wide Web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation,* 43:209-226.

Barsalou, L.W., (1999). Perceptual Symbol systems. *Behavioral and Brain Sciences,* 22: 577-660.

Bentivogli, L., Forner, P., Magnini, B., Pianta, E. (2004). Revising WordNet Domains Hierarchy: Semantics, Coverage, and Balancing. In Proceedings of COLING 2004 Workshop on "Multilingual Linguistic Resources", Geneva, Switzerland, August 28, 2004, 101-108.

Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, bhp055.

Binder, J.R., Westbury, C.F., McKiernan, K.A. Possing, E.T., Medler, D.A., (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*. 17:905-917.

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, &*

Computers, 30, 188-198.

Chang, K. M., Mitchell, T., Just, M. A. (2011). Quantitative modeling of the neural representation of objects: How semantic feature norms can account for fMRI activation, *NeuroImage* 56 (2011) 716–727.

Fellbaum, C., (1998, ed.). WordNet: An Electronic Lexical Database. Cambridge, MA: MIT Press.

Friederici, A.D., Ruschemeyer S-A, Hahne A., Fiebach, C.J., (2003). The role of left inferior frontal gyrus and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. *Cereb Cortex, 13:170-177.*

Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L., (2002). Sweetening Ontologies with DOLCE. In A. Gómez-Pérez, V.R. Benjamins (eds.) Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web, 13th International Conference, EKAW 2002, Siguenza, Spain, October 1-4, 2002, Springer Verlag, pp. 166-181

Gentner, D., (1981). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. *In S. A. Kuczaj, editor, Language development:* 2:301-334. Erlbaum, Hillsdale, NJ.

Grossman, M., Koenig, P., DeVita, C., Glosser, G., Alsop, D., Detre, J., (2002). The neural basis for category specific knowledge: An fMRI study. *Neuroimage*, 16:936-948.

Hampton, J., (1981). An investigation of the nature of abstract concepts. *Memory & Cognition,* 9(2):149-156.

Haxby, J.V., Gobbini, M.I., Furey, M.L, Ishai, A., Schouten, J.L., Pietrini, P., (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293:2425-2430.

Kriegeskorte, N., Mur. M., Bandettini, P., (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2:4.

Landauer, T.K., Dumais, S.T., (1997). A solution to Plato's problem: The latent semantic analysis, theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211-240.

Landauer, T.K., Foltz, P.W., Laham, D., (1998). An introduction to latent semantic analysis. *Discourse Processes*, 27:303-310.

Lenat, D. and Guha, R. V., (1990). Building large Knowledge-based systems: Representation and inference in the Cyc Project. *Addision-Weslely.*

Lund, K., Burgess, C., (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instrumentation and Computers*, 28: 203-208.

McRae, K., Cree, G.S., Seidenberg, M.S., McNorgan, C., (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, and computers*, 37(4):547-559.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K.M., Malave, V. L., Mason R. A., and Just., M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns, *Science*, 320, 1191-1195. DOI: 10.1126/science.1152876

Murphy, B., Baroni, M., Poesio, M. (2009). EEG Responds to Conceptual Stimuli and Corpus Semantics. Proceedings of ACL/EMNLP 2009.

Murphy, B., Poesio. M, Bovolo, F., Bruzzone, L., Dalponte, M., Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, 117, 12-22.

Olivetti, E., Greiner, S., & Avesani, P. (2012). Testing multiclass pattern discrimination. In IEEE International Workshop on Pattern Recognition in NeuroImaging (PRNI). 57-60 DOI:10.1109/PRNI.2012.14

Pianta, E., Bentivogli, L., Girardi., C., (2002). MultiWordNet: developing an aligned multilingual database" pdf document. In Proceedings of the First International Conference on Global WordNet, Mysore, India, January 21-25, 2002.

Wang, J., Conder, J.A., Blitzer, D.N., Shinkareva, S.V. (2010). Neural representation of abstract and concrete concepts: A meta-analysis of neuroimaging studies. *Human Brain Mapping*, 31:1459-1468.

Wang, J., Baucom, L.B., Shinkareva, S.V. (2012). Decoding abstract and concrete concept representations based on single-trial fMRI data. *Human Brain Mapping*, DOI: 10.1002/hbm.21498

Warrington, E.K. & Shallice, T., (1984). Category specific semantic impairments. *Brain,* 107(3):829-853.

Wiemer-Hastings, K., Xu, X. (2005). Content differences for abstract and concrete concepts. *Cognitive Science*, 29:719-736.