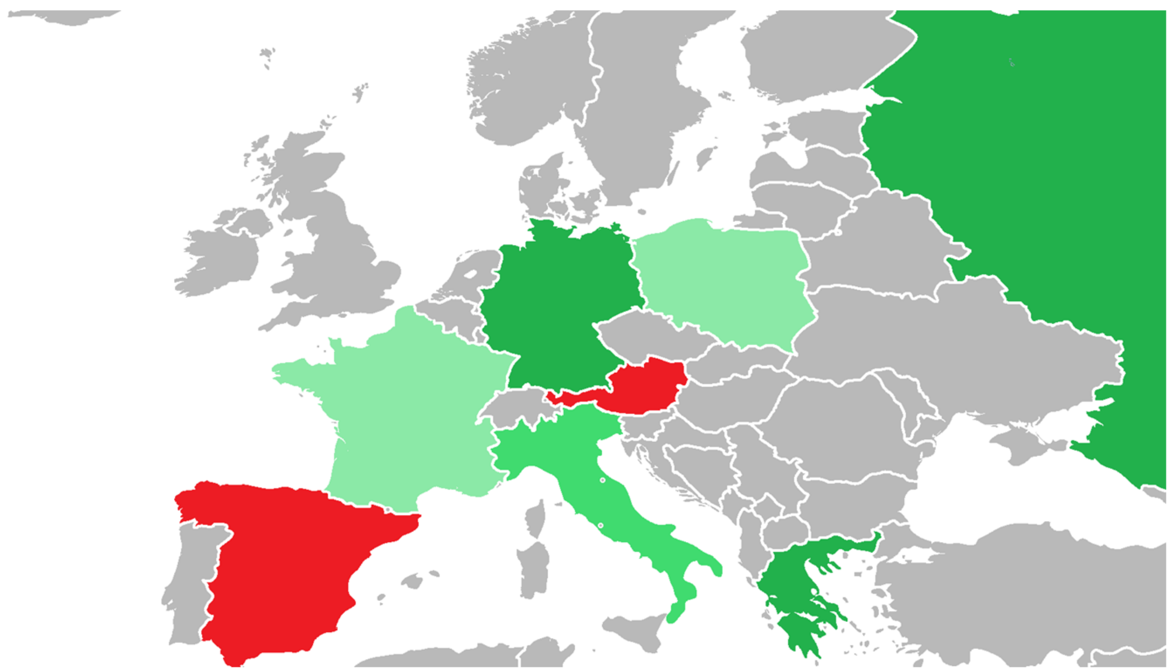
Evidence of Semantic Processing in the Brain

A symposium on the occasion of Patrick Suppes' 90th birthday
March 9---10---11, 2012
Stanford University



Colleen Crangle, Marcos Perreau-Guimaraes, Patrick Suppes

EXPERIMENTAL SETUP

A computer presented forty eight spoken sentences to each participant, in 10 randomized blocks with all 48 sentences in each block. So there were 480 trials.

The sentences were about the geography of Europe. Half were true, half false, half positive, half negative.

The capital of Italy is Paris.	F
Paris is not east of Berlin.	T
Spain is west of Russia.	T

Participants were asked to determine the **truth or falsity** of each statement while EEG recordings were made.

X ϵ {**Berlin, London, Moscow, Paris, Rome, Warsaw, Madrid, Vienna, Athens**)
Y ϵ {**France, Germany, Italy, Poland, Russia, Austria, Greece, Spain**)
W ϵ {the capital, the largest city}
Z ϵ {**north, south, east, west**}

W of Y is [not] X

X is [not] W of Y

X is [not] Z of X   Y is [not] Z of Y

Two fundamental claims established in the 2009 paper:

Suppes, P., M. Perreau-Guimaraes, D. K. Wong (2009). Partial Orders of Similarity Differences Invariant Between EEG-Recorded Brain and Perceptual Representations of Language. *Neural Computation* 21, 3228–3269.

1. **When brain wave recordings are time-locked to the presentation of words in a sentential context, segments of those brain waves (after appropriate transformations) can be identified as representations of those words.**

   These representations have approximate invariance across language speakers (test participants) and across a variety of sentential contexts.

2. **Significant structural similarities can be found between language constituents (phonemes, for example) and the brain waves identified as their representations.**

What I show today

- Structural similarities that can be found between the brain data and the language data **at the semantic level.**

- New methods to measure the **strength** of these structural similarities. These measures allow us to **assess the relatively better fit with the brain data of one semantic model over another**.

What I show today

- Structural similarities that can be found between the brain data and the language data **at the semantic level.**

- New methods to measure the **strength** of these structural similarities. These measures allow us to **assess the relatively better fit with the brain data of one semantic model over another**.

Three semantic models

WordNet, a lexical database of English organized around sets of cognitive synonyms that can be strongly restricted by **context**

Latent Semantic Analysis (LSA), a statistical method to extract measures of word similarity from selected sets of documents such as novels, newspaper articles, textbooks. Restricted **contexts**.

A probability-based method that extracts word relations from the internet -- with no restriction on **context**

WORDNET

Paris

(n) **Paris#1**, City of Light#1, French capital#1, capital of France#1 (the capital and largest city of France; and international center of culture and commerce)

(n) **Paris#2**, genus Paris#1 (sometimes placed in subfamily Trilliaceae)

(n) **Paris#3** ((Greek mythology) the prince of Troy who abducted Helen from her husband Menelaus and provoked the Trojan War)

(n) **Paris#4** (a town in northeastern Texas)

Multiple word senses

Interrelated sets of synonyms

Directed Acyclic Graph

LATENT SEMANTIC ANALYSIS (LSA)

A statistical technique for extracting from large bodies of text a measure of how similar two words are to each other -- in terms of how similar their contexts of use are – in a particular set of documents.

General college-level novels

Psychology textbook

…

GOOGLE-BASED NORMALIZED SIMILARITY SCORE (NSS)

If two words X and Y never occur together on the same web page, but do occur separately, the normalized Google distance (NGD) between them approaches infinity.

If X and Y always occur together, their NGD is zero.

Intuitively, NGD is an **estimate of the symmetric conditional probability of co-occurrence of the words X and Y.**

NGD → NSS.

LSA and NSS (Google) provide straightforward measures of similarity between words.

For WORDNET, several different measures of similarity – path length between synonym sets, for example – have been devised. I used 5 different measures of similarity and took the average.

	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	1.000	0.396	0.466	0.106	0.103	0.076	0.078	0.322	0.299	0.303
Moscow	0.396	1.000	0.393	0.095	0.094	0.062	0.070	0.286	0.281	0.288
Paris	0.466	0.393	1.000	0.106	0.104	0.074	0.077	0.327	0.308	0.307
north	0.106	0.095	0.106	1.000	0.228	0.179	0.210	0.123	0.132	0.111
south	0.103	0.094	0.104	0.228	1.000	0.172	0.212	0.115	0.107	0.109
east	0.076	0.062	0.074	0.179	0.172	1.000	0.216	0.093	0.080	0.077
west	0.078	0.070	0.077	0.210	0.212	0.216	1.000	0.087	0.082	0.083
Germany	0.322	0.286	0.327	0.123	0.115	0.093	0.087	1.000	0.589	0.409
Poland	0.299	0.281	0.308	0.132	0.107	0.080	0.082	0.589	1.000	0.403
Russia	0.303	0.288	0.307	0.111	0.109	0.077	0.083	0.409	0.403	1.000

Semantic similarity matrix derived from WordNet for the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*} using senses relevant to the geography of Europe and five measures of similarity: wup (path length), lin and jcn (information content), and gv and pgv (vector space measures)

Structural models for brain and language data

The first step is to analyze the brain data using a statistical model that predicts to which class a brain data sample belongs.

We used a 5-fold linear discriminant model, with principal component analysis for blind source separation, to classify the segments of data obtained from the individual trials.

For our 10 geography words ***{London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}*** 640 data samples are classified into 10 classes.

More generally, the $\mathbf{T}$ brain data samples $\mathbf{s_1, s_2, \ldots, s_T}$ are classified into the $\mathbf{N}$ classes $\boldsymbol{\omega_1, \omega_2, \ldots, \omega_N}$ of the finite set $\mathbf{A}$. In this process, we do not directly compare the samples to each other to judge their similarity; rather we use a set of $\mathbf{N}$ prototypes, and these are what are designated by $\boldsymbol{\omega_1, \omega_2, \ldots, \omega_N}$.

If test sample $\mathbf{s_i}$ is classified as $\boldsymbol{\omega_i}$ then $\mathbf{s_i}$ and $\boldsymbol{\omega_i}$ have a minimal similarity difference compared to the other possible classifications.

Classify **640** <u>brain data</u> samples $\mathbf{s_1, s_2, \ldots, s_{640}}$ into **10** classes $\boldsymbol{\omega_1, \omega_2, \ldots, \omega_{10}}$ of the finite set $\mathbf{A}$.

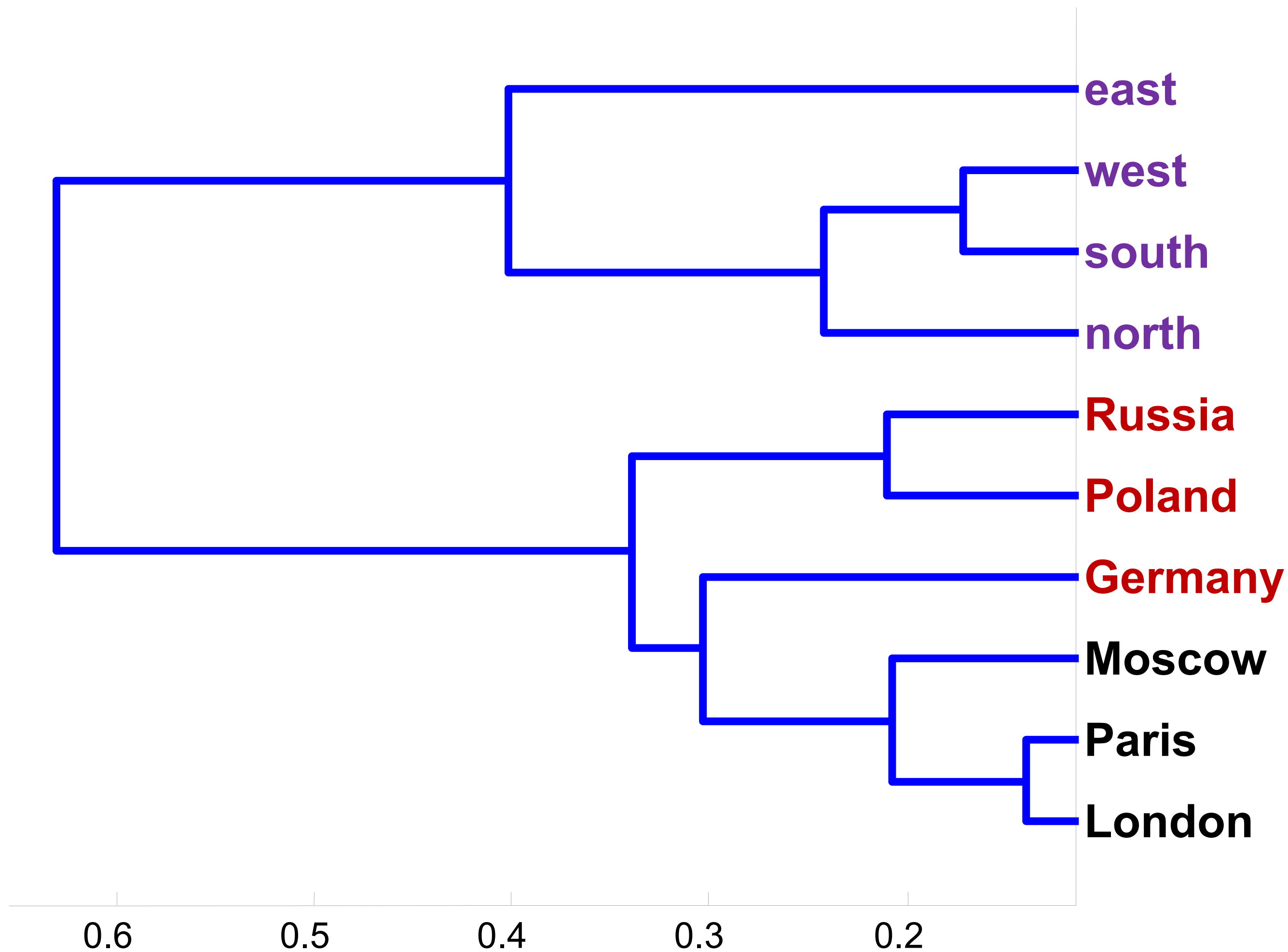
$\mathbf{M} = (\mathbf{m_{iq}})$ is the **confusion matrix** for a given classification, where $\mathbf{m_{iq}}$ is the number of test samples from class $\boldsymbol{\omega_i}$ classified as belonging to class $\boldsymbol{\omega_q}$.

	London	***Moscow***	***Paris***	***north***	***south***	***east***	***west***	***Germany***	***Poland***	***Russia***
London	8	14	11	7	6	3	3	10	8	10
Moscow	8	24	14	6	2	6	4	4	7	15
Paris	6	18	12	4	3	5	8	6	11	7
north	4	2	5	11	9	7	10	1	8	3
south	1	4	3	14	14	9	11	4	7	3
east	4	3	5	9	12	12	7	1	4	3
west	4	3	2	12	13	11	10	2	8	5
Germany	2	2	3	2	2	1	0	9	11	8
Poland	2	3	4	0	4	1	4	9	9	4
Russia	7	7	5	1	5	1	1	8	4	11

The relative frequencies $\mathbf{m_{iq}}$ / $\sum_q \mathbf{m_{iq}}$ are an N-by-N estimate for the conditional probability densities – designated by the matrix $\mathbf{P} = (\mathbf{p_{iq}})$ – that a randomly chosen test sample from class $\boldsymbol{\omega_i}$ will be classified as belonging to class $\boldsymbol{\omega_q}$.

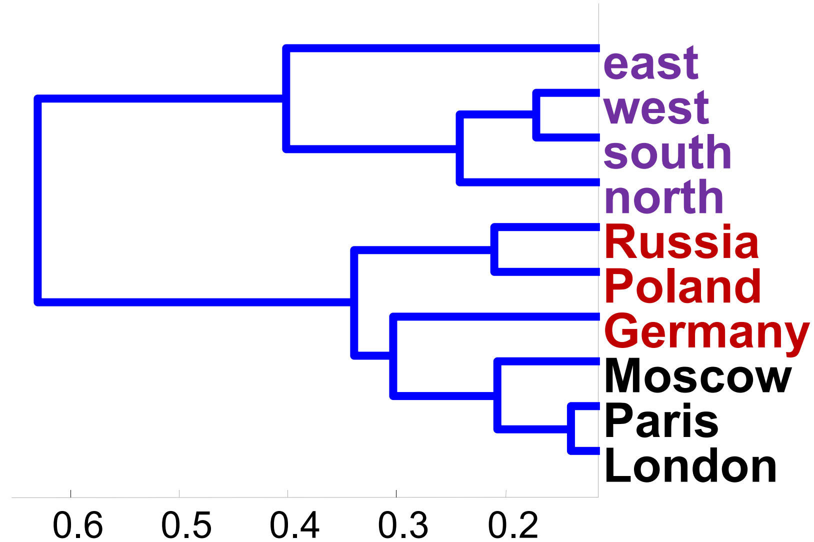
	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	0.250	0.163	0.138	0.075	0.038	0.050	0.063	0.088	0.075	0.063
Moscow	0.144	0.333	0.133	0.033	0.011	0.056	0.056	0.089	0.067	0.078
Paris	0.175	0.188	0.125	0.038	0.038	0.038	0.050	0.100	0.138	0.113
north	0.067	0.017	0.017	0.283	0.167	0.150	0.100	0.100	0.067	0.033
south	0.014	0.014	0.071	0.114	0.271	0.171	0.157	0.029	0.086	0.071
east	0.050	0.000	0.000	0.133	0.150	0.383	0.150	0.050	0.000	0.083
west	0.057	0.043	0.000	0.171	0.114	0.143	0.229	0.086	0.100	0.057
Germany	0.075	0.175	0.100	0.050	0.125	0.025	0.050	0.275	0.025	0.100
Poland	0.150	0.100	0.125	0.050	0.050	0.000	0.025	0.150	0.250	0.100
Russia	0.160	0.040	0.080	0.040	0.100	0.020	0.120	0.100	0.080	0.260

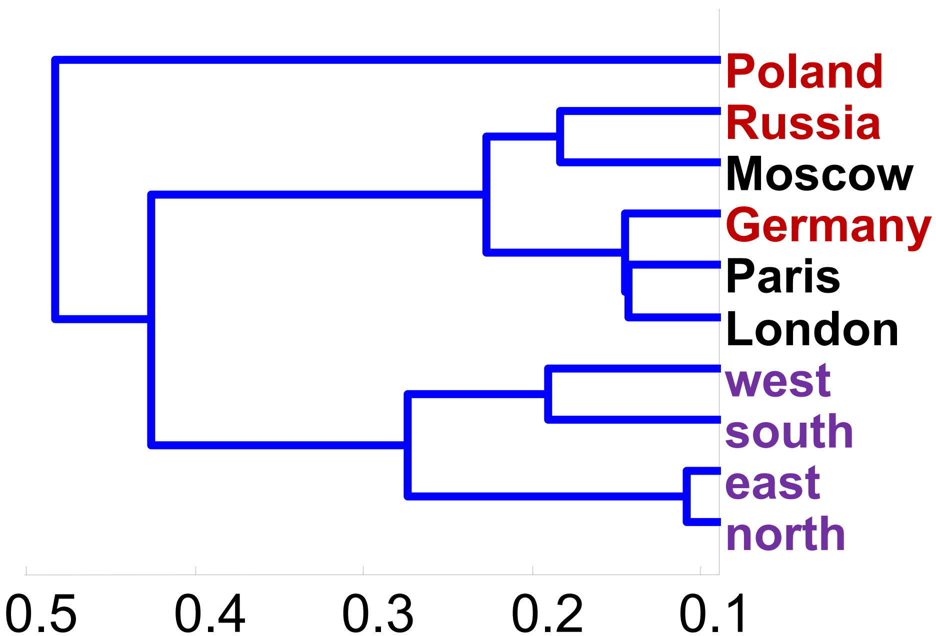
Conditional probability density estimates from the classification of brain wave data for {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}



Hierarchical cluster tree (similarity tree) computed from the conditional probability density estimates for the classification of 640 brain wave samples for {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}

Every single-trial classification produces its own conditional probability density estimates giving rise to its own similarity tree ...





Another hierarchical cluster tree (similarity tree) computed from the conditional probability density estimates for the classification of 640 brain wave samples for {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}

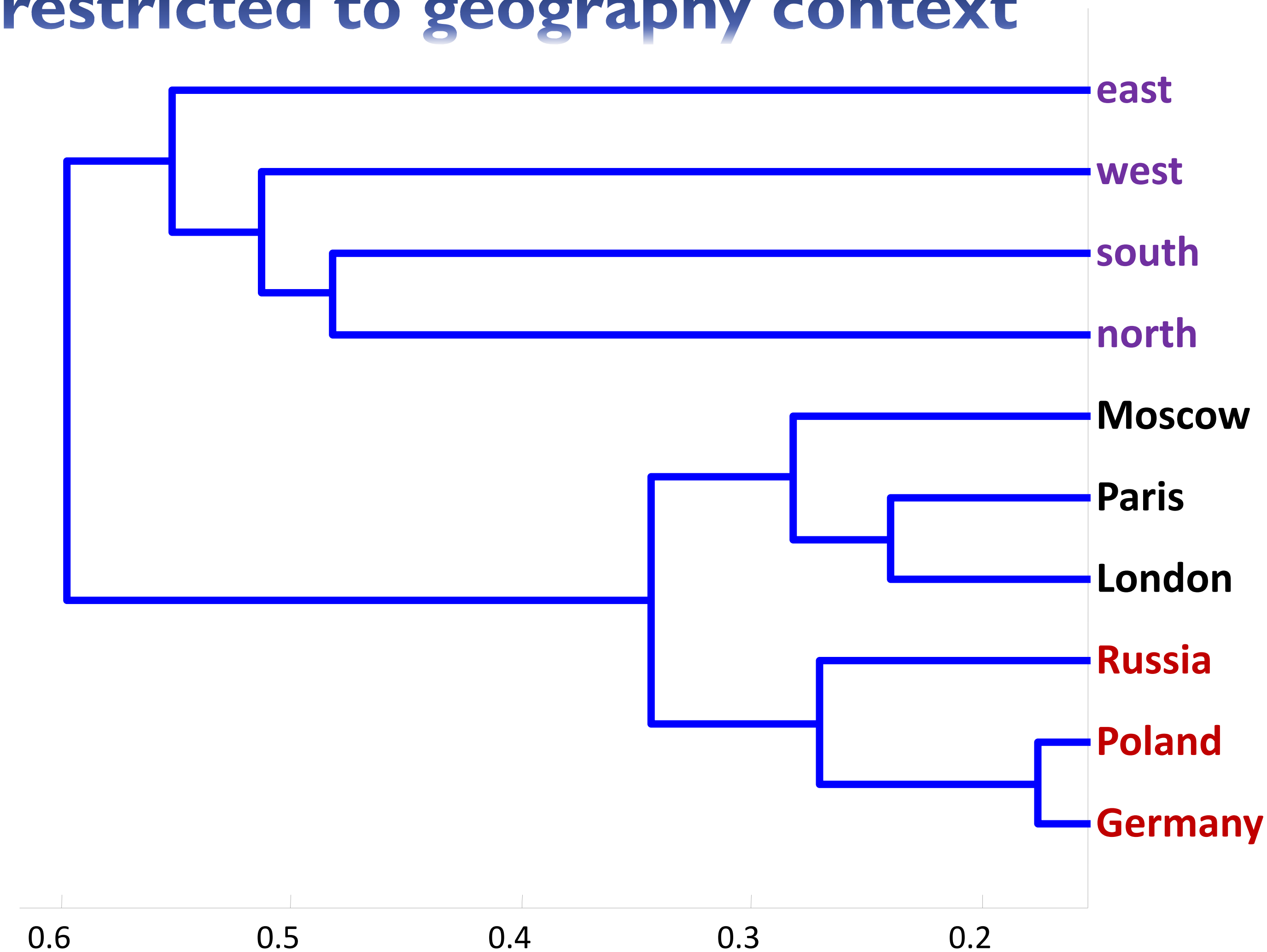
Back to semantics …

Shown earlier

	London	Moscow	Paris	north	south	east	west	Germany	Poland	Russia
London	1.000	0.396	0.466	0.106	0.103	0.076	0.078	0.322	0.299	0.303
Moscow	0.396	1.000	0.393	0.095	0.094	0.062	0.070	0.286	0.281	0.288
Paris	0.466	0.393	1.000	0.106	0.104	0.074	0.077	0.327	0.308	0.307
north	0.106	0.095	0.106	1.000	0.228	0.179	0.210	0.123	0.132	0.111
south	0.103	0.094	0.104	0.228	1.000	0.172	0.212	0.115	0.107	0.109
east	0.076	0.062	0.074	0.179	0.172	1.000	0.216	0.093	0.080	0.077
west	0.078	0.070	0.077	0.210	0.212	0.216	1.000	0.087	0.082	0.083
Germany	0.322	0.286	0.327	0.123	0.115	0.093	0.087	1.000	0.589	0.409
Poland	0.299	0.281	0.308	0.132	0.107	0.080	0.082	0.589	1.000	0.403
Russia	0.303	0.288	0.307	0.111	0.109	0.077	0.083	0.409	0.403	1.000

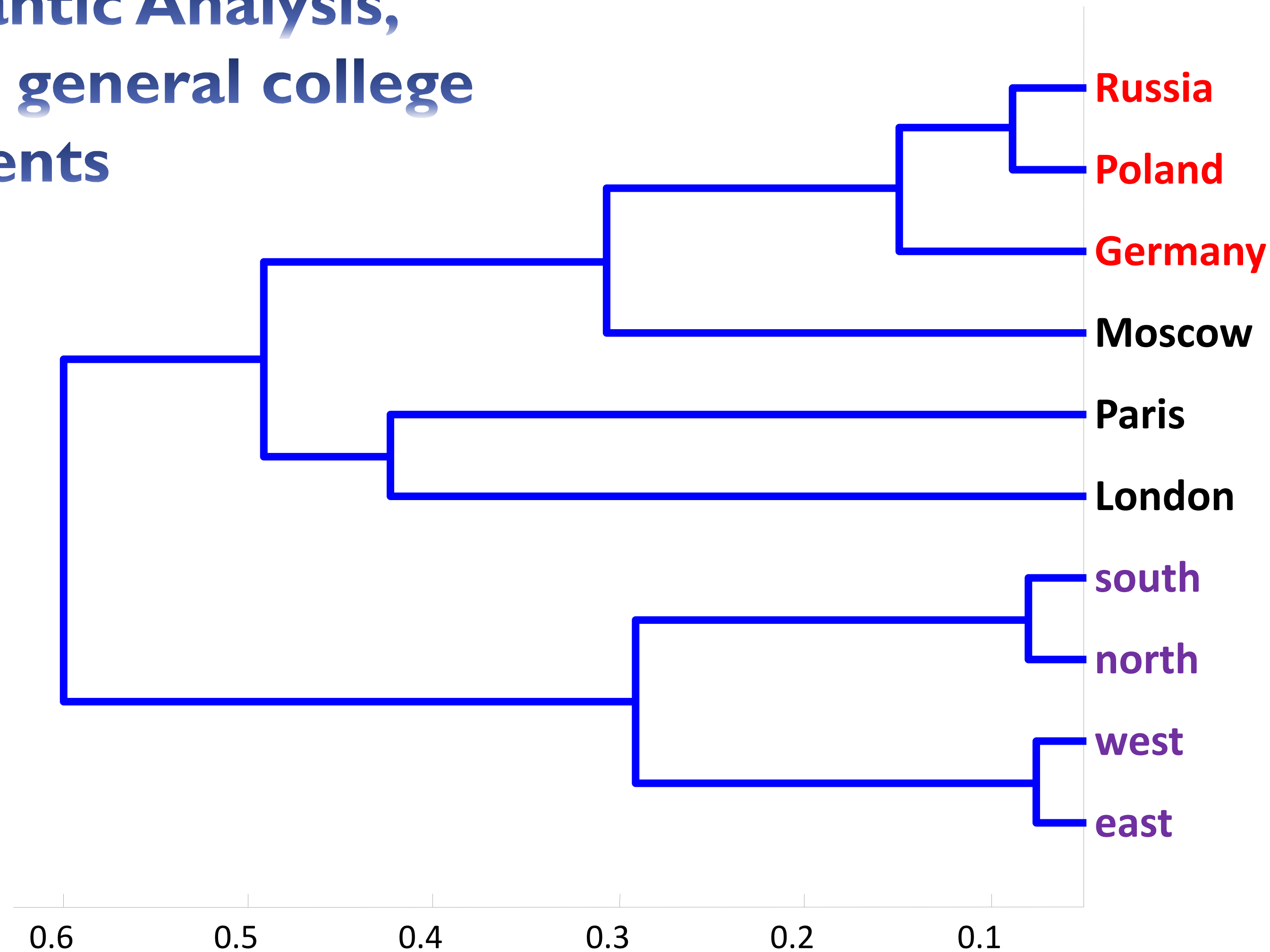
Semantic similarity matrix derived from WordNet for the set of words {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*} <u>using senses relevant to the geography of Europe</u> and five measures of similarity: wup (path length), lin and jcn (information content), and gv and pgv (vector space measures)

WordNet, restricted to geography context



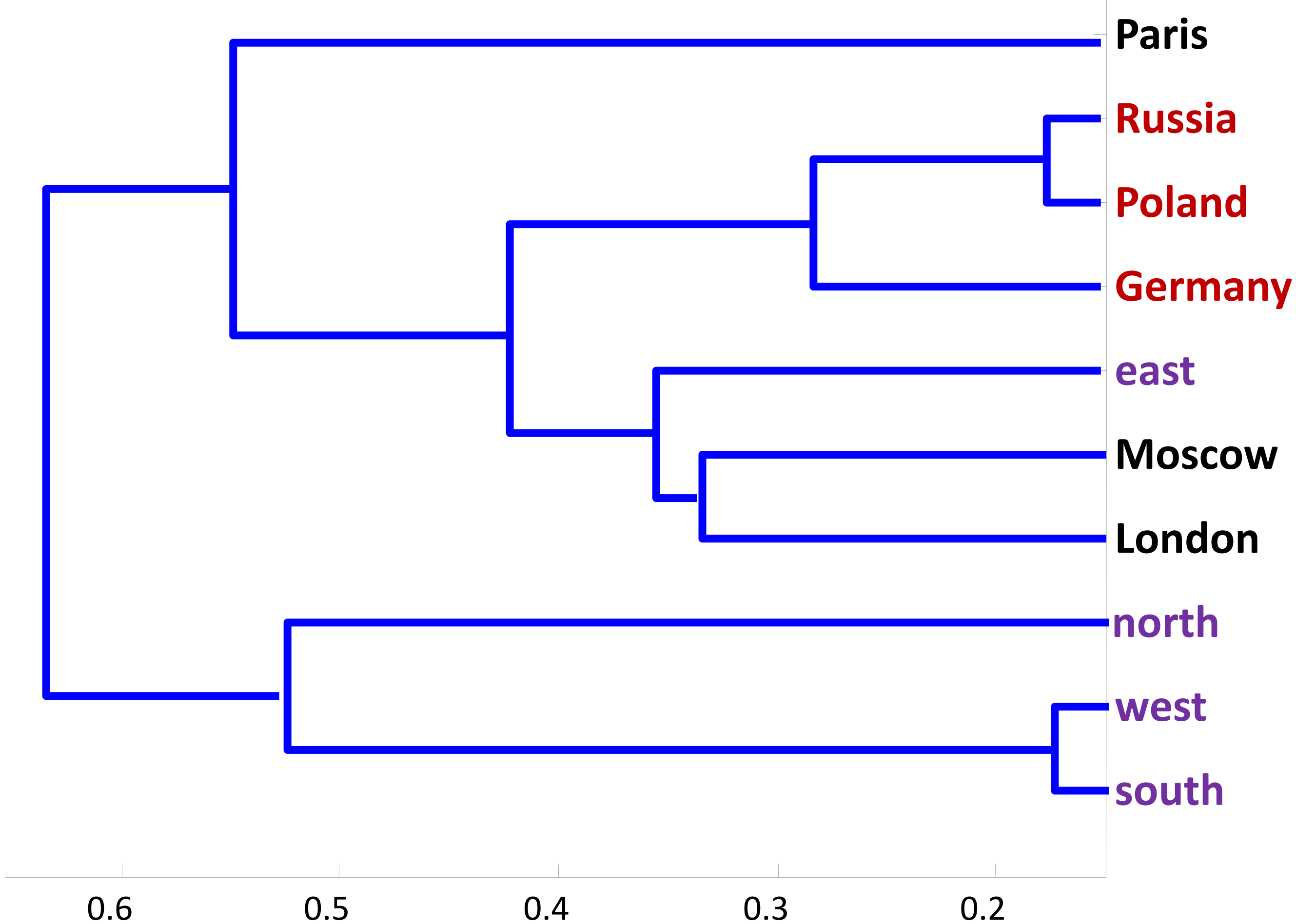
Hierarchical cluster tree computed from pairwise WordNet-based semantic similarity scores for {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*} restricted to senses related to the geography of Europe

Latent Semantic Analysis, restricted to general college level documents



Hierarchical cluster tree computed from the pair-wise Latent Semantic Analysis (LSA) scores of similarity for {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*} based on ~ 38,000 college-level texts (novels, newspaper articles,…)

Google-based Normalized Similarity Scores



Hierarchical cluster tree computed from the contextually unrestricted Google-based normalized similarity scores (NSS) for {*London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia*}

Some of the similarity trees show remarkable congruence between the brain and semantic data

Where exactly does that congruence lie?

Can we devise a quantitative measure of the nature and strength of that congruence?

For each word ω we compute from the conditional probability density estimates a ternary relation **R'** such that **R'(** ω, ω_1, ω_2 **)** if and only if with respect to word ω the conditional probability for word ω_1 is smaller than the conditional probability for word ω_2 , that is if and only if ω_1's similarity difference with ω is smaller than ω_2's similarity difference with ω.

R' is an ordinal relation of similarity differences, a partial order that is irreflexive, asymmetric, and transitive.

And we do exactly the same for the similarity matrices of the semantic models.

For each word ω we compute from the semantic similarity matrix a ternary relation **R''** such that **R'' (ω', ω'$_1$, ω'$_2$)** if and only the similarity difference of ω'$_1$ with ω' is smaller than the similarity difference of ω'$_2$ with ω', that is, ω'$_1$ is more similar to ω' than is ω'$_2$.

R'' is an ordinal relation of similarity differences, a partial order that is irreflexive, asymmetric, and transitive.

Following the approach described in Suppes (1974) for the axiomatization of the theory of differences in utility preference or the theory of differences in psychological intensity

Suppes, P. (1974) The axiomatic method in the empirical sciences. L. Henkin et al. (Eds.), *Proceedings of the Tarski Symposium, Proceedings of Symposia in Pure Mathematics*, **25**. Providence, RI: American Mathematical Society, pp. 465-479.

The relational structure **(A, R')** constructed from **R'** and the finite set **A** of classes $\boldsymbol{\omega_1, \omega_2, \ldots, \omega_N}$ together with the <u>**N** partial orders</u> constructed from the N-by-N estimate for the conditional probability densities

⟷

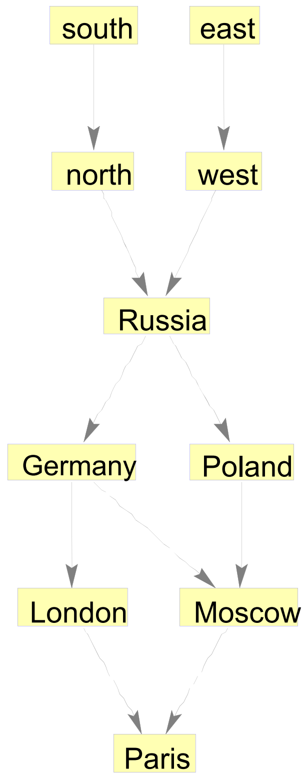
The relational structure **(A', R'')** constructed from **R''** and the finite set **A'** of classes $\boldsymbol{\omega'_1, \omega'_2, \ldots, \omega'_N}$ together with the <u>**N** partial orders</u> constructed from the N-by-N similarity matrix

Brain data

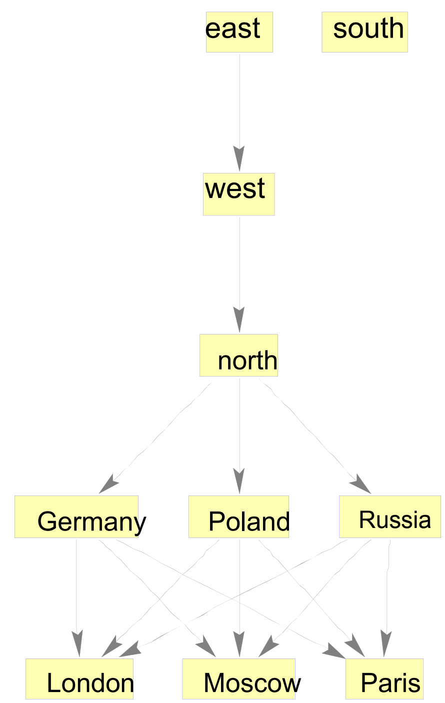
Language data

For each $\boldsymbol{\omega_1}$ we **compare the partial order of the brain data with the partial order of the language data**, using Spearman's rank correlation coefficient, which we interpret in the usual way to determine if we have a statistically significant correlation or not.

For those instances in which the brain and language partial orders are significantly correlated, *we find the partial order that is invariant with respect to the brain and language data*

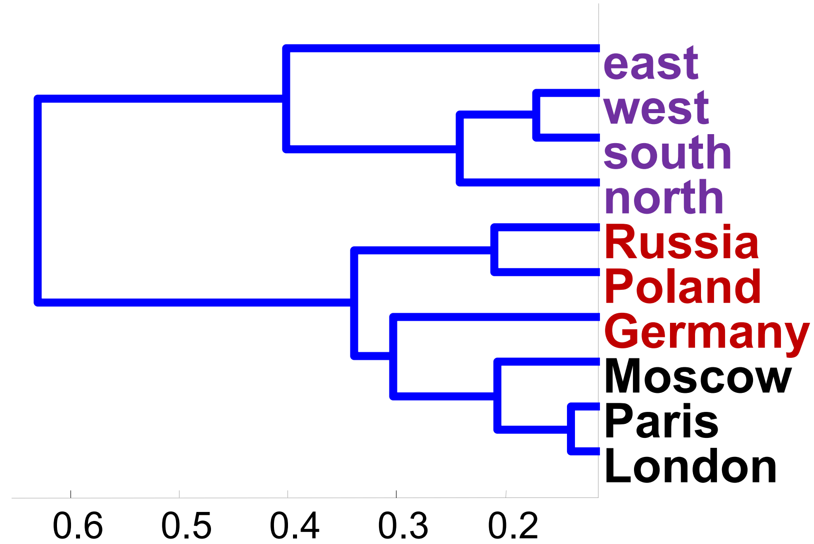


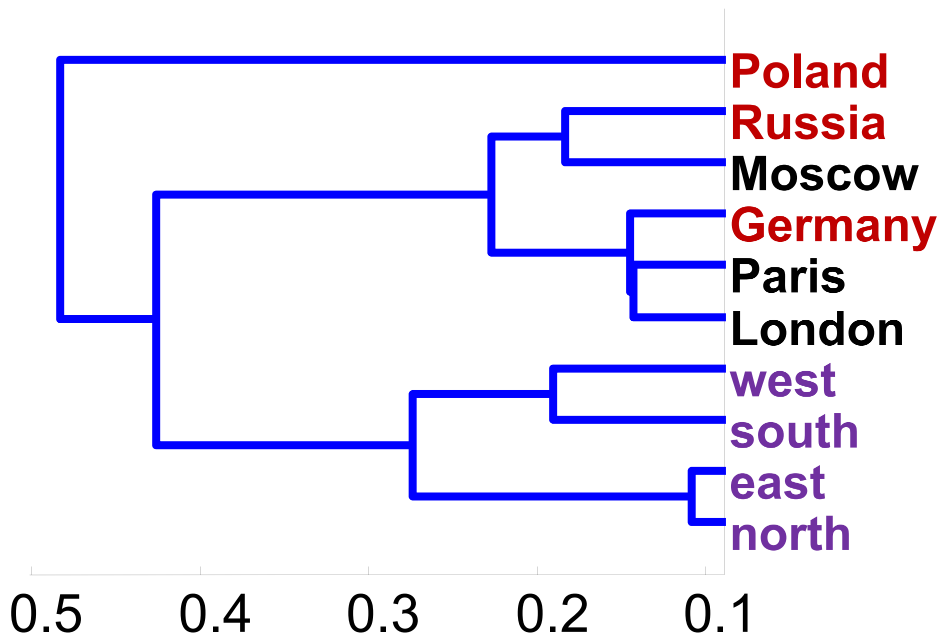
Significant Invariance - Paris - Spearman 0.88 (p=6.6795e-00



Significant Invariance - Paris - Spearman 0.90 (p=3.8716e-004)

As pointed out earlier, every single-trial classification produces its own conditional probability density estimates giving rise to its own similarity tree …





Another hierarchical cluster tree (similarity tree) computed from the conditional probability density estimates for the classification of 640 brain wave samples for *{London, Moscow, Paris, north, south, east, west, Germany, Poland, Russia}*

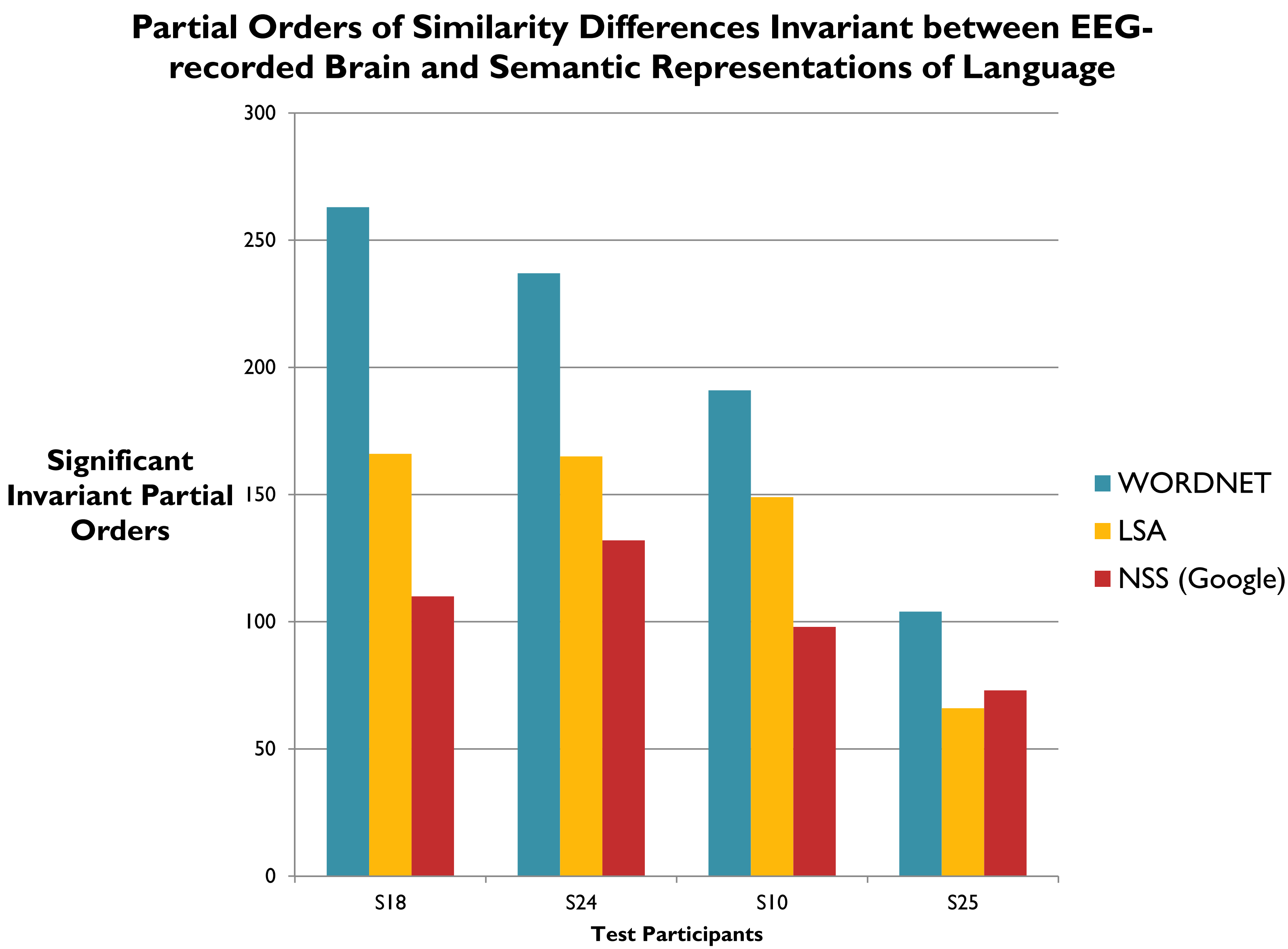
We therefore perform a series of computations.

1. We compute M (=30) single-trial classifications of the data (640 data samples for our 10 words) using random resampling with replacement.
2. For each classification we find for each word the partial orders of the brain and language data that are significantly correlated.
3. For each of these highly correlated partial order pairs, we find the partial order invariant with respect to both.

We do this for

1. **the WordNet data**
2. **the LSA data**
3. **the Google-based NSS data**

and plot the results.


Partial Orders of Similarity Differences Invariant between EEG-recorded Brain and Semantic Representations of Language
Significant Invariant Partial Orders
300
250
200
150
100
50
0
S18
S24
S10
S25
Test Participants
WORDNET
LSA
NSS (Google)

The original conclusion I drew from these results is that …

The brain data reveal evidence of context-sensitive processing. The number of statistically significant structural similarities found between the brain and language data is greatest when the language model is strongly restricted to be contextually relevant [WORDNET senses restricted to the geography of Europe].

For a language model that is less subject to contextual restrictions [LSA], the brain and language data exhibit fewer statistically significant structural similarities.

For a language model that is subject to no contextual restrictions [Google-based NSS], the brain and language data exhibit even fewer statistically significant structural similarities.

Do these results instead tell us something about the semantic models? That WordNet is a better model of some aspect of language processing than LSA? A better model of how the lexicon is organized in the mind? A better model of semantic knowledge?