

A potential surface underlying meaning?

Darányi, S.^{*}, Wittek, P.,^{*†}, Konstantinidis, K.^{**},
Papadopoulos, S.^{**}

^{*} Swedish School of Library and Information Science, University of Borås,
Sweden

[†] Institute of Photonic Sciences (ICFO), Spain

^{**} Centre for Research and Technology Hellas (CERTH), Greece

October 6, 2015

Yandex School of Data Analysis Conference
Machine Learning: Prospects and Applications

What is meaning?

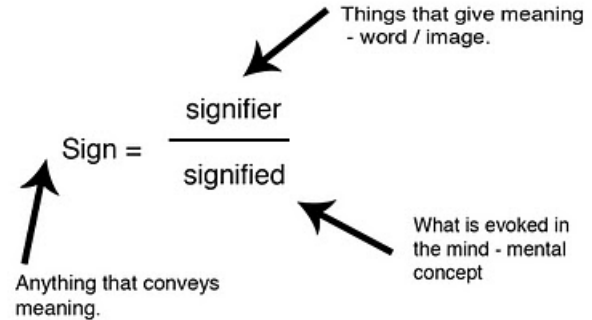
- A universal goal of information seeking, but we don't know for sure in the natural science sense.
 - Scalable applications handling the problem increasingly available, many theories of word and sentence meaning, but **no single unified model** after 2000 years.
 - Different theories capture different aspects.
 - “Ghost in the machine”: given a vector representation of semantic content, where does meaning hide in the numbers?
- Relevant theories include:
 - The contextual approach (distributional hypothesis [Harris 1970], see also Wittgenstein [“The meaning of words lies in their use”, *Phil. Inv.* (1953) 80, 109] and Firth [“You shall know a word by the company it keeps” (1957), 11]).
 - The referential approach: a meaning of a word is defined in a dictionary, ontology etc. (Frege 1948).
 - The relational approach: sense relations between word pairs help to structure the vocabulary (Lyons 1968, Fellbaum 1998).
 - The field approach: related concepts map to regions in semantic space (Trier 1934).
- IR and TC models tap into the *distributional hypothesis* to associate meaning with form.

What's wrong with that?

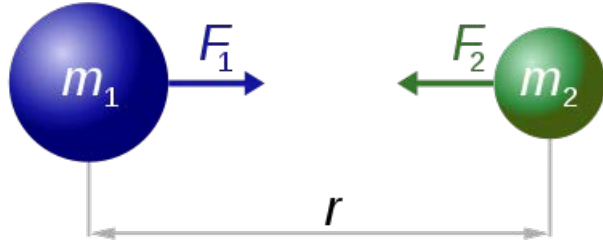
- Nothing is wrong with the distributional hypothesis, but its interpretation is incomplete.
 - Vector spaces built on it have almost everything in Newton's 2nd law, $F = ma$, save for m (cf. Salton 1968).
- Assume that machine learning (ML) algorithms utilizing gradient descent work for word semantics because of a so far not studied possibility:
 - Algorithm efficiency in text categorization is always subject to best match with human judgment as a yardstick.
 - Therefore identifying local and global minima with learnables such as concepts presupposes the existence of a **vocabulary** represented **as a surface** against which fitness comparison can take place.
 - A simple example to illustrate this idea is a *potential surface underlying gravitation as a force*, which is the subject of Newton's above law.
- But if distance between localized items is a given in vector space, whereas mass – or its equivalent, charge in Coulomb's law – are unknown in semantics:
 - On the one hand, how to circumvent this?
 - On the other hand, apparently, weighting schemes provide one with a quantitative aspect of meaning captured by term frequencies only, while the qualitative aspect that distinguishes two words is currently not addressed.
- This leads to the working hypothesis that **physical fields could be a useful metaphor to study word and sentence semantics**.
 - As surfaces are studied in classical mechanics, we want to test CM for the modelling of indexing terminology change.

So, just for now...

- Consider meaning as **located content** in vector space:
 - The linguistic sign unites a concept with its sound-image (Saussure), or form with substance (St. Augustine)...
 - ...like a particle “unites” location and energy.



A link with the RBF kernel



$$F_1 = F_2 = G \frac{m_1 \times m_2}{r^2}$$

$$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$$

- $F = Gm_1m_2/r^2$ is **Newton's law of universal gravitation**.
- If we consider similarity as an attractive force, the RBF kernel can create the impression of gravitation:
 - It has an distance based inverse square decay component of interaction strength,
 - Plus a scalar scaling factor for the interaction,
 - But no term masses.

Similarity, force, potential, field

- In [classical mechanics](#), the **gravitational potential** at a location is equal to the [work](#) ([energy](#) transferred) per unit mass that would be done by the force of [gravity](#) if an object were moved from its location in space to a fixed reference location.
- To look at kernels measuring **similarity as an attractive force** implies the idea of an underlying **semantic potential**, in tandem with the study of **semantics as a vector field**.
- A vector field can be constructed from term co-occurrences as a source of semantics by the Emergent Self-Organizing Maps (ESOM) algorithm (Ultsch & Moerchen 2005):
 - Already emerged vs. still emerging content are clearly distinguishable, much in line with Aristotle's *Metaphysics*, as shown on different text collections.
 - Suitable to study actual vs. potential locations of content, i.e. the dynamics of evolving semantics.
- Gut feeling: a model completed by “term mass” or “term charge” would enable the computation of the specific work equivalent of sentences or documents, and that replacing semantics by other modalities, vector fields of more general symbolic content can exist.
- We take the first steps toward this goal as follows.

“Forces” in language, “energy” in ML: a clarification

- Language change implies measurable “forces”. Examples:
 - Cross-textual cohesion and coherence (White 2002):
 - Strong nuclear force (*strongest*) = word uninterruptability (binds morphemes into words).
 - Electromagnetism (*less strong*) = grammar (binds words into sentences).
 - Weak nuclear force (*less strong*) = texture/cohesion; coherence (binds sentences into texts).
 - Gravity (*weakest*) = intercohesion (binds texts into literatures).
 - Co-occurrence and collocation.
 - Attraction (synonymy, hypernymy, holonymy) vs. repulsion (antonymy, esp. graded antonymy) by sense relations.
 - Lexical attraction and repulsion (Beeferman *et al.* 1997); valence theory and dependency grammars (Tesnière 1959).
- For categorization as minimization of an energy function:
 - Greek *energeia* (Aristotle *Metaphysics* ix.8 1050a22), here: “work capacity, work content” [in a structure].
 - In ML, the notion is used for differentiable nonlinear functions, e.g. gradient descent based algorithms identify concepts as local vs. global minima.
 - Mathematical energy examples:
 - Signal energy in calculations, devoid of physical content (e.g. Park 2003).
 - “Signals that arise from strictly mathematical processes and have no apparent physical equivalent are commonly considered to represent some form of mathematical energy” (Bruce 2007)
 - Local density of values in mathematical object: “Energy of a (part of a) vector is calculated by summing up the squares of the values in the (part of the) vector” (Wang & Wang 2001)

A trick and a tool

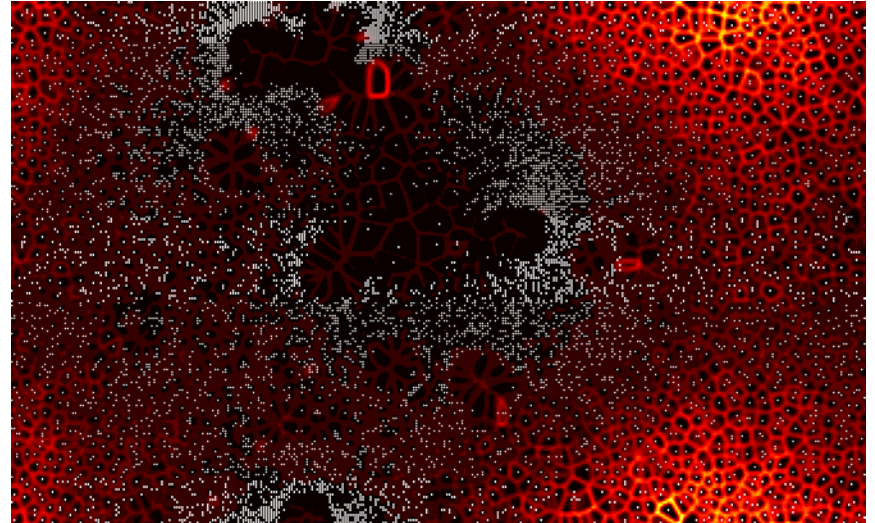
- The trick to find the missing link is to introduce an **index term mass**:
 - Such as PageRank to compute F .
 - F measures **node similarity AND node popularity/influence/importance**.
 - It models interaction between language and social behaviour.
 - Probably could be extended to different measures of semantic relatedness (MSR).
- Next, we need a tool to compute term positions and dislocations over time in a vector field to characterize evolving semantics by F .

From vector space to vector field semantics

Semantics as tectonics

- Tool for the experiment: [Somoclu](#), a massively parallel software implementing ESOM
- Models the vocabulary of the test dataset as a vector field:
 - Approximately 5 grid nodes per index term to interpolate a field from positions, originally to study lexical gaps in word distributional patterns (Wittek *et al.* 2014).
 - Here, adapted to study lexical cohesion on forces.
 - Lexical cohesion, also called collocation: when two sentence elements share a lexical field.
 - White dots are BMUs for terms, red hot vs. black indicates difference in dynamics, yellow boundaries hint at high probability of term displacement, i.e. tensions causing splits in the structure.

Vector field of Reuters-21578 index terms

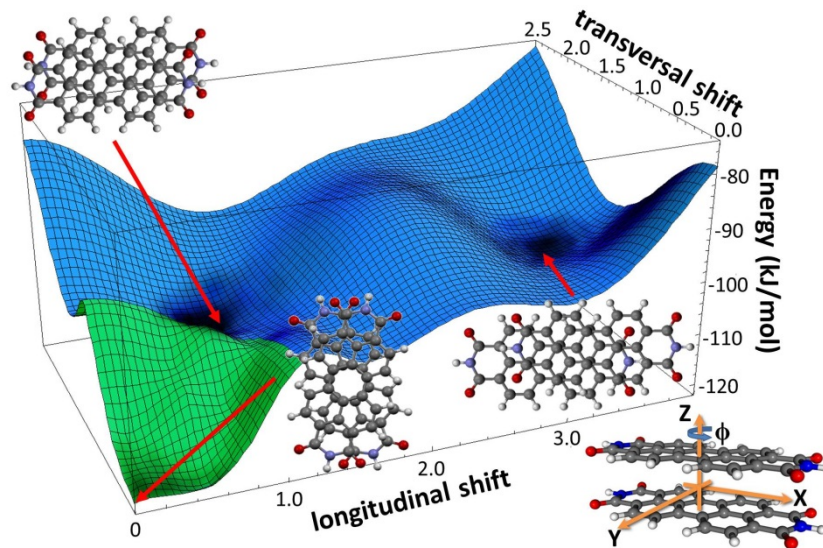


What are we looking for?

Two models in blueprint:

- Consider semantic content as a conservative field:
 - Plan A: with a non-polar force and a combination of interaction and external potential modelled on gravitation.
 - Plan B: with a polar force and a Lennard-Jones-like potential, modelled on chemistry.

Something like [this](#): structure-dependent energy



Plan A: Experiment design, step 1

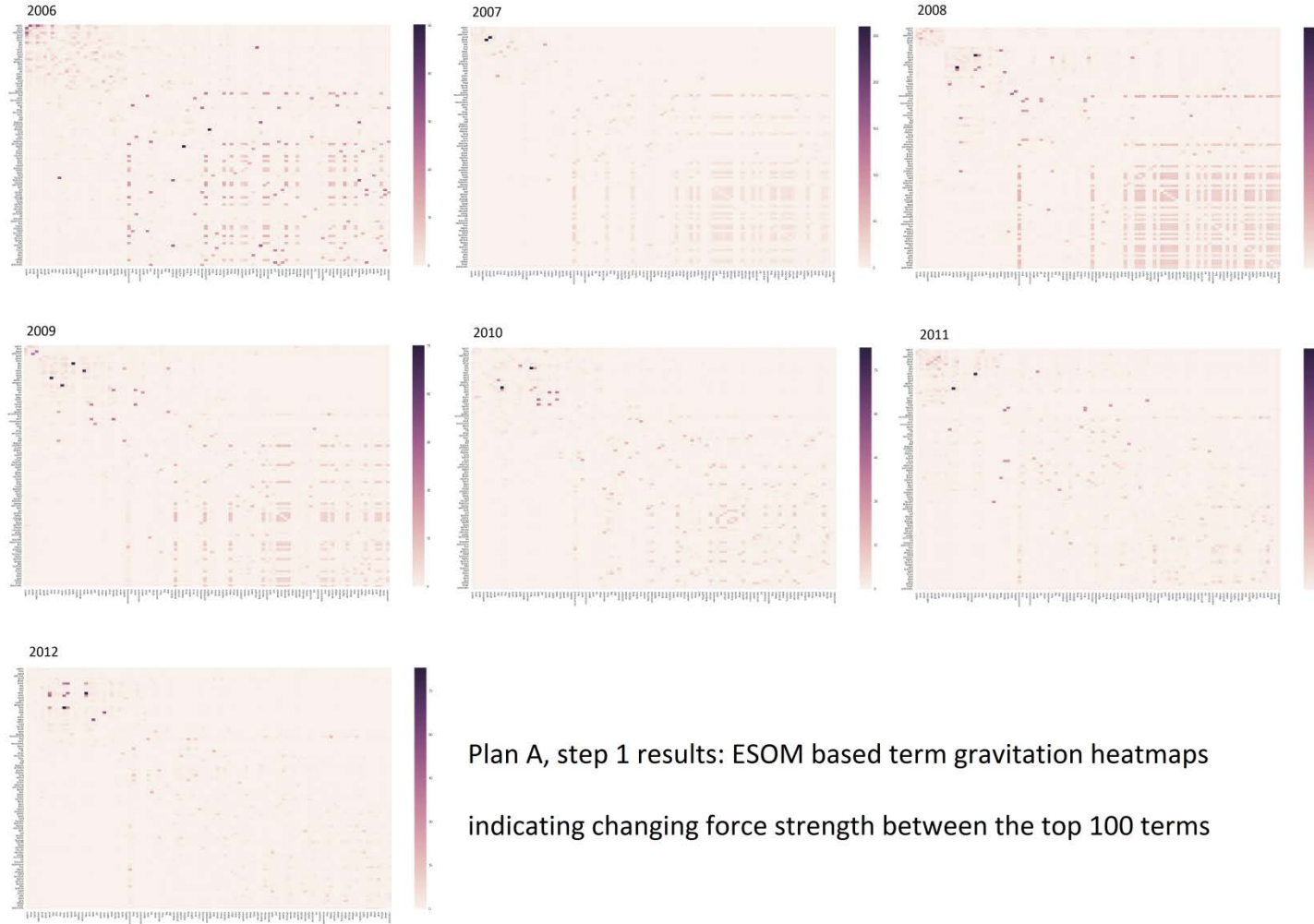
- Corpus: Amazon watch reviews dataset from [SNAP](#) (2006-2012).
- 68.356 reviews by 17.000 unique terms over 7 years.
- 4801 terms showing up at least once in every timeslot: more consistent dataset by pruning.
- Out of these 4801, the top 100 occurring over the complete period were selected. Each term had a PR (NPR) value assigned to it for every year (even if it was a dangling node). To find the most consistently central terms, the sum of [reciprocal ranks](#) over the years was calculated and the first 100 terms were retained.
- Using the period-specific term distance vs. term mass (PR/NPR) matrices, the gravity matrices were computed, normalized by the max for NPR vs. sum for PR.
 - (Following a +1 increment to all values in order to avoid dividing by zero).
- Out of these, the gravitation of the top 100 over the 4801 was extracted to yield a "field" by interpolation.

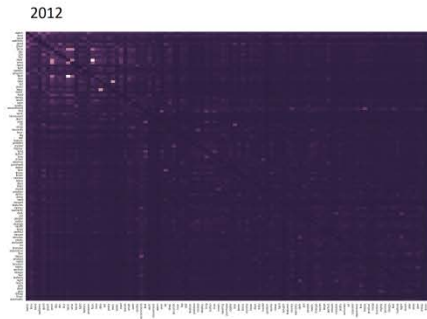
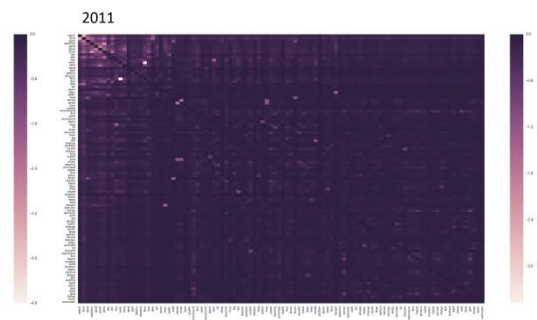
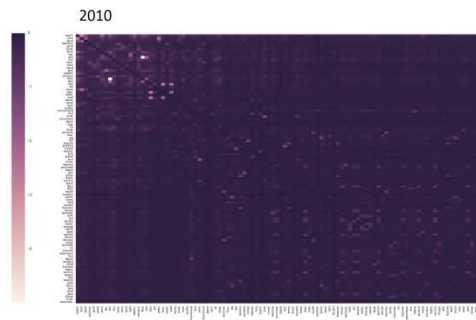
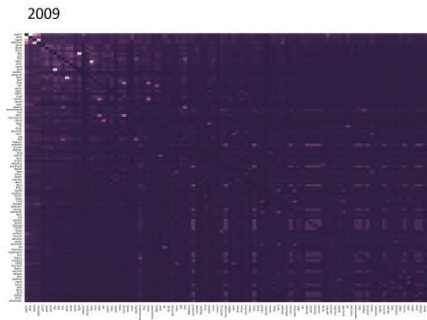
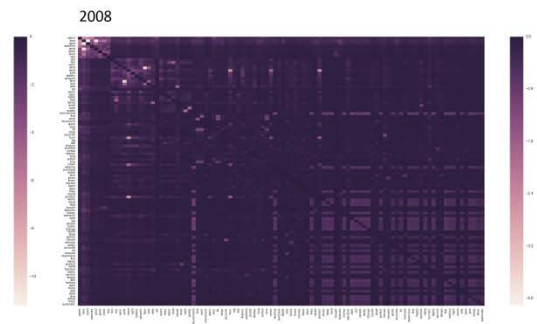
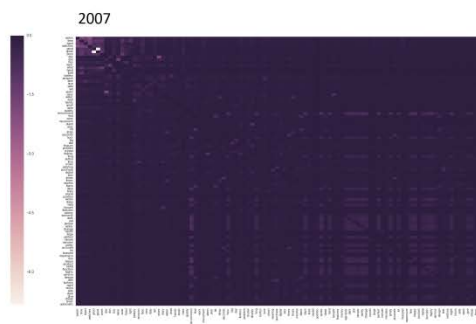
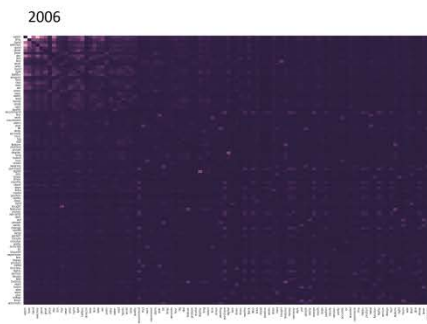
Plan A: Experiment design, step 2 – work in progress

- Embedding the 100 top terms in 2D with toroid ESOMs.
- From the ESOM results, recalculate velocities and accelerations combined with the PR (NPR) values, to yield the **total kinetic energy** (KE) and **interaction potential** (IP) for every term in each period.
 - These are components of the Hamiltonian of the system, $H = T + V$.

---- WE ARE THIS FAR WITH THE EXPERIMENT ----

- For a closed system, KE and IP may fluctuate between the time periods, because the total energy of the system also includes the energy of the external potential (EP).
- Importance:
 - The extent of the fluctuation of the KE + IP could indicate the strength of interaction between the particles (terms) and the EP.
- Calculate the value of the EP for each point of the grid from this single value (the fluctuation of the total energy) by a reasonable heuristic.
- Plot the EP.





Plan A, step 1 results cont'd.: ESOM based term potential heatmaps
indicating changing gravitational potential between the top 100 terms

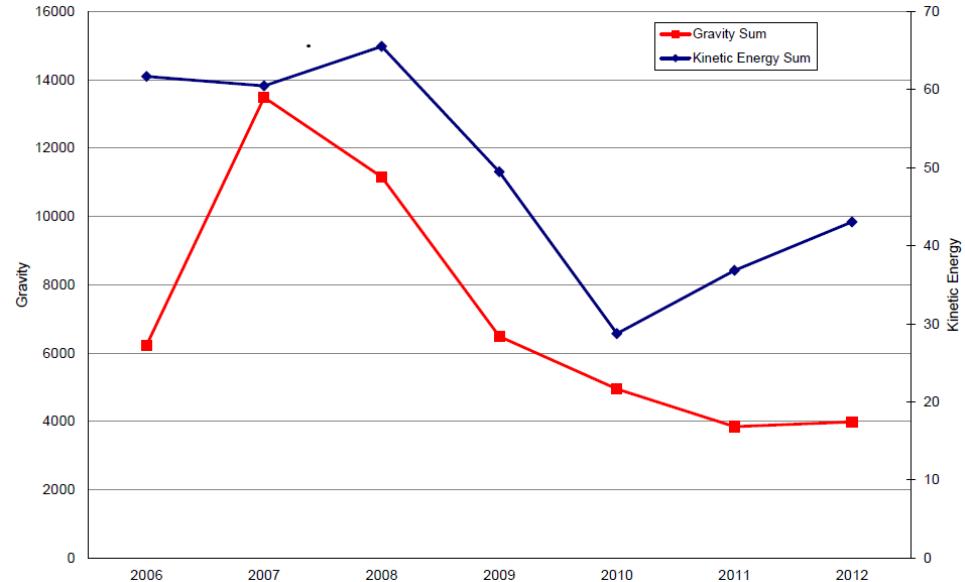
Parallel changes in term gravitation and potential indicate topic shifts

Strongly bonded term pairs (incomplete manual evaluation)

2006	2007	2008	2009	2010	2011	2012
time-watch	great-good	face-wrist	light-face	face-don	face-wear	face-wear
watch-watches		nice-wrist	light-day	date-read	small-hands	easy-date
buy-bought		nice-face	face-wear	date-easy	year-buy	face-wrist
digital-display		quality-recommend	watches-band	easy-read	love-hands	wrist-price
feature-features			hands-work		love-small	wear-price
features-buttons						

Plan A, step 2 results: hint at an external potential

- We assumed a closed system:
 - For the time being, only humans generate language, and only on this planet.
 - Pruning closed down the dataset.
- That means that the total energy of the system does not change over time.
- The total energy has **three major parts**: sum of individual kinetic energy of particles + sum of pairwise interaction energy of particles + sum of energy with the interaction of an external potential.
- Now we only have the first two components, but the energy is not constant, which means there is a missing third component, which is the impact of the external potential.
- That component is necessary to compute the so far hidden potential surface.

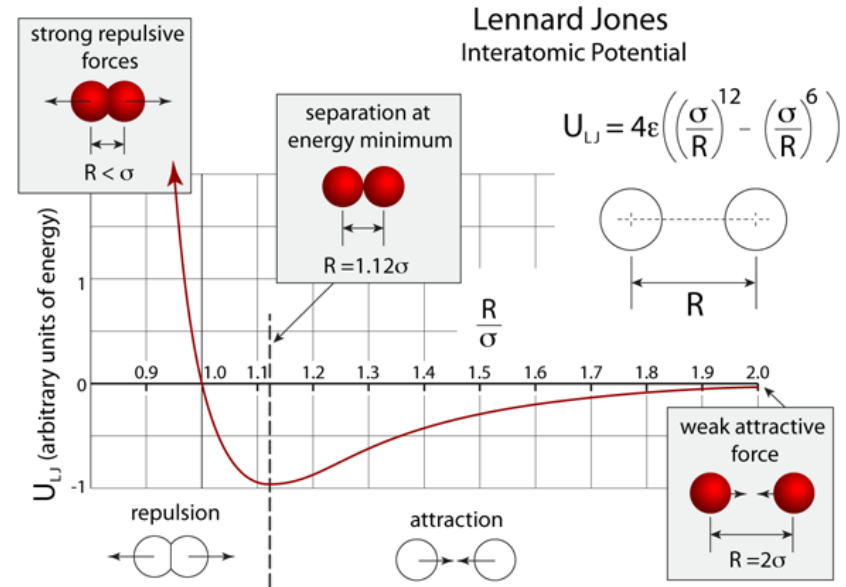


Plan A, step 2 results – more work in progress

Plan A, 2nd (alternative) scenario:

- [Nonpolar] Gravitation alone is not enough to model evolving semantics, hence the gravitational potential falls short of what we want. Reasoning:
 - *Sim* measures the overlap between two binary vectors in terms of identity/difference in position values.
 - Attraction expresses similarity only, to represent difference we also need repulsion.
 - It is their balance that keeps terms in position.
 - **As if** an external factor prevented the vocabulary from collapsing onto a single word with a million meanings.
 - This external factor can be expressed by an external potential as part of the Hamiltonian.
 - Their interplay is similar to the L-J potential but not necessarily identical with it.

Lennard-Jones potential



Summary

- The hypothesis of a physical equation matching the data passed statistical testing.
- Total KE + IP fluctuates over the period, hints at the option of an external potential.

Future research: the neuroscience link

Fallisard, B. (2011). A thought experiment reconciling neuroscience and psychoanalysis. *Journal of Psychology* 105, 201-206.

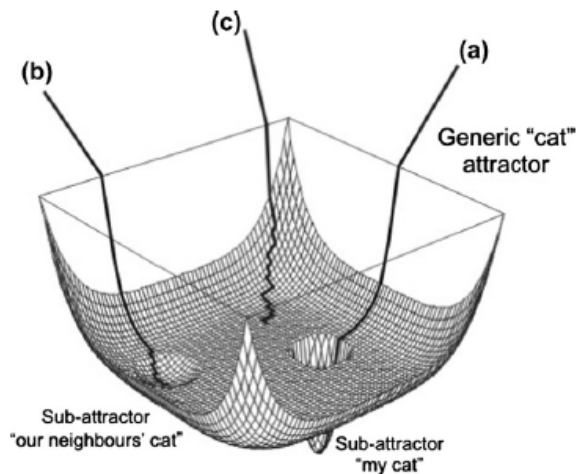


Fig. 1. Geometrical representation of brain processing. Each node of the grid is a given configuration of a neural network. Height is the level of energy of each configuration. The network becomes stable when it reaches a minimum of energy (attractor).

Continuous to discrete

- In neuroscience, continuous vehicle of percepts modeled as a grid by NN.
- In linguistics/semiotics, continuous conceptual field mapped onto discrete lexical field by content sampling.
- Mappings yield actual positions for lexemes, transitions between concepts remain in potentiality called lexical gaps.
- Can be modelled by grid nodes for actual lexemes vs. interpolated values for transitions.
- Concepts as attractors in a conceptual field, with lexemes as lexical attractors in its respective lexical field mapping.
- This view is supported by neurosemantics where concrete noun representations are stored in a spectral fashion (Just *et al* 2010).

“Meaning is interaction between mental state and word as cue” (Elman 2004)

Selected bibliography

- Beeferman, D., Berger, A., Lafferty, J. 1997. A model of lexical attraction and repulsion. In: *Proceedings of ACL-97*, 35th Annual Meeting of the Association for Computational Linguistics, Madrid, Spain, ACL, Morristown, NJ, USA 373-380.
- Bruce, E. 2001. *Biomedical signal processing and signal modeling*. New York: Wiley.
- Elman, J.L. 2004. An alternative view of the mental lexicon. *Trends in Cognitive Sciences* 8(7), 301-306.
- Fellbaum, C. /Ed./ 1998. *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Firth, J.R. 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.
- Frege, G. 1948. Sense and reference. *The Philosophical Review* 57(3), 209-230.
- Harris, Z. 1970. Distributional structure. In Harris, Z. /Ed./: *Papers in structural and transformational linguistics*. New York: Humanities Press. 775-794.
- Just, M.A., Cherkassky, V.L., Aryal, S., Mitchell, T.M. (2010). A neurosemantic theory of concrete noun representation based on the underlying brain codes.. *PLoS ONE* 5(1), e8622.
- LeCun, Y., Chopra, S., Hadsell, R. 2006. A tutorial on energy-based learning. In: *Predicting Structured Data*. 1-59.
- Lyons, J. 1968. *Introduction to theoretical linguistics*. New York: Cambridge University Press.
- Osgood, C.E., Suci, G.J., Tannenbaum, P.H. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.
- Park, L. 2003. *Spectral Based Information Retrieval*. PhD thesis, University of Melbourne.
- Salton, G. 1968. *Automatic information organization and retrieval*. New York: McGraw-Hill.
- Tesnière, L. 1959. *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Trier, J. 1934. Das sprachliche Feld. *Neue Jahrbücher für Wissenschaft und Jugendbildung* 10, 428-449.
- Ultsch, A., & Moerchen, F. 2005. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. *Technical Report Dept. of Mathematics and Computer Science, University of Marburg, Germany*, No. 46.
- Wang, C., Wang, X. 2001. Indexing very high-dimensional sparse and quasi-sparse vectors for similarity searches. *The VLDB Journal* 9(4), 344-361.
- White, H. 2002. Cross-textual cohesion and coherence. In: *Proceedings of the Workshop on Discourse Architectures: The Design and Analysis of Computer-Mediated Conversation*, Minneapolis, MN, USA.
- Wittek, P., Darányi, S., Liu, Y-H. 2014. A Vector Field Approach to Lexical Semantics. In: *Proceedings of 8th International Conference on Quantum Interaction*, Filzbach, Switzerland. June 30 - July 3, 2014.
- Wittek, P., Darányi, S., Kontopoulos, E., Mysiadis, T., Kompatsiaris, I. 2015. *Monitoring Term Drift Based on Semantic Consistency in an Evolving Vector Field*. At <http://arxiv.org/abs/1502.01753>
- Wittgenstein, L. 1953. *Philosophical investigations*. Oxford: Blackwell.