Connecting the Dots: Mass, Energy, Word Meaning, and Particle-Wave Duality

Sándor Darányi and Peter Wittek

Swedish School of Library and Information Science
University of Borås
Allégatan 1, 50190 Borås, Sweden

Abstract. With insight from linguistics that degrees of text cohesion are similar to forces in physics, and the frequent use of the energy concept in text categorization by machine learning, we consider the applicability of particle-wave duality to semantic content inherent in index terms. Wave-like interpretations go back to the regional nature of such content, utilizing functions for its representation, whereas content as a particle can be conveniently modelled by position vectors. Interestingly, wave packets behave like particles, lending credibility to the duality hypothesis. We show in a classical mechanics framework how metaphorical term mass can be computed.

1   Introduction

The general idea to use physics, prominently both classical and quantum mechanics, to model phenomena crucial to managing society has lately made an interesting debut among significant research questions. To do so, the calculus being the same as in quantum mechanics, the trick is to switch the probability type from the variant pertinent in the subatomic domain to the other one which works in the macroworld [1]. By this, phenomena like financial market evolution (deterministic and stochastic models of markets), language evolution, or digital repositories handling term meaning fluctuations become available for extensive and application-oriented, testable research with implementable results.

At the same time, currently this intellectual tightrope walking is mostly theory development, with expected immediate practical implications. If one can show on new use cases such as e.g. media, health, and part text, part signal based science data that quantum-like modelling works, and yields at least as good or better results than the benchmark, then one will have to ask for the reason of this success. One explanation we have at this point is the metaphorically "energetic" nature of both language [2] and learning (decision making) [3]. Since we perceive this "energetic" nature based on macroworld observations also underpinned

by microworld calculation methodology, the phenomenon must be underlying both classical mechanics (CM) and quantum mechanics (QM).

With this caveat, as a next phase in an ongoing thought experiment, below we will first cite arguments from linguistics to compare kinds of coherence in language to types of forces in physics, and suggest the field concept as a possible unifying metaphor to represent word and sentence semantics both in the CM and QM frames of thought. Our first observation cites views calling for a regionally rather than exactly located kind of semantic content to be considered [4]. Secondly, mathematical objects with physical equivalents, such as sinusoids standing for e.g. electromagnetic waves and wavelets for impulses (outbursts), can be used to model the above regional distribution of semantic content [5,6], also demonstrated for the visible light part of the spectrum [7] – as if, in a sense, terms and documents had a wave nature. Thirdly, we will postulate that the same semantic content can be conceived both as a particle and a wave-packet, arguing for this quasi-dual nature as an essential property enabling new kinds of experiments.

The topic is heating up because of increasing interest in temporal dynamics [8–11] and its anticipated connection with the Hamiltonian, a typically quantum interaction (QI) consideration. As proposed earlier, in both CM and QM, it is the Hamiltonian which describes the energy stored in a system, and in order to approach it, finding a way to compute term "mass" is the key.

This paper is organized as follows: Section 2 outlines the concept of energy in machine learning, Section 3 interprets the Hamiltonian on the constituents of a term-document matrix and in Section 4 we consider the duality of semantic content representation. Section 5 is the discussion of some immediate implications of our observations, with Section 6 offering our brief conclusions.

2 Energy in machine learning

The metaphoric use of physics is based on the urge to find better models of text classification (TC) and information retrieval (IR) by means of machine learning (ML). We start with arguments from linguistics to compare kinds of coherence in language to types of force in physics.

As White suggests, linguistics, like physics, has four binding forces [12]:

– The strong nuclear force, which is the strongest "glue" in physics, corresponds to word uninterruptability (binding morphemes into words);

- Electromagnetism, which is less strong, corresponds to grammar and binds words into sentences;
- The weak nuclear force, being even less strong, compares to texture or cohesion (also called coherence), binding sentences into texts;
- Finally gravity as the weakest force acts like intercohesion or intercoherence which binds texts into literatures (i.e. documents into collections or databases).

Mainstream linguistics traditionally deals with Forces 1 and 2, while discourse analysis and text linguistics are particularly concerned with Force 3. The field most identified with the study of Force 4 is information science [12–14]. As the concept of force implies, referring here to attraction, it takes energy to keep things together, therefore the energy doing so is stored in agglomerations of observables of different kinds in different magnitudes, and can be released from such structures. A notable difference between physical and lingustic systems is that extracting work content, i.e. "energy" from symbols by reading or copying them does not annihilate symbolic content, however.

Looking now at the same problem from a different angle, in the above and related efforts, energy inherent in all four types can be the model of:

- A Type 2, i.e. electromagnetism-like attractive-repulsive binding force such as lexical attraction, a.k.a syntactic word affinity [15], also called sentence cohesion, such as by modelling dependency grammar by mutual information [16]. Once pointwise mutual information replaces mutual information, the nickname of the effect is "infomagnetism" [17]. In a TC and/or IR setting, a similar phenomenon is term dependence based on their co-occurrence;
- Decision making, such as in a classification process, both in a supervised and unsupervised manner;
- Information representation, such as conceiving documents as wave interference patterns [7, 18].

Again from a different angle, the energy concept used in experiments of the above types can be mathematical or physical. For mathematical energy, at least three concepts are current:

- Signal energy in calculations, devoid of physical content (e.g. [6]. A typical consideration runs like this: "Signals that arise from strictly mathematical processes and have no apparent physical equivalent are commonly considered to represent some form of mathematical energy" [19];

- Loss functions in ML. These model the cost of a classification decision as an energy minimizing process;
- Local density of values within a 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" [20]

For physical energy as a model of content or processes, the applicability of the metaphor in ML pertains to all kinds of media, images included. With a focus mainly on clustering, its range is already strikingly broad, spanning the electromagnetic force [21], gravity [22, 23], spin [24], waves [25], wavelets [26], and wave functions [27, 28]. At the same time, the utilization of the energy concept in ML goes back to the use of potentials. In the examples cited above, there are two kinds thereof, Coulomb potential vs. gravitational potential, so that decision making (classification, categorization) is minimum or maximum seeking by gradient descent or ascent on a hypersurface, constructed from statistics describing the event space. Whereas gravitational force assumes energy from the mass of particles (i.e., documents) in a cluster, Coulomb potential presupposes the dipole nature of entities such as belonging to vs. not belonging to a class.

We have to mention here the inherent eigen conjecture in latent semantic methods, namely that terms have "mass", i.e. word meaning behaves as if it had an energetic nature [2]. This conjecture is now reinforced by the explicit consideration of mass in the kinetic part of dynamic quantum clustering (DQC) as well [28].

3 The role of the Hamiltonian in evolving document collections

Taking into consideration that $H = T + V$ is the Hamiltonian equation we want to interpret, where H is the Hamiltonian operator, T is the kinetic energy and V is the potential energy of a system, respectively, we argue that $AA^{\mathrm{T}} = H$, that is, we treat the term co-occurrence matrix as the description of the total energy of the system. Thereby we also assume that our system is a conservative one. The same assumption was made by DQC [27–29].

Any update of AA^{T} results in an $A'A'^{\mathrm{T}}$ state with its corresponding V' potential energy, whereas the difference between any two consecutive V' goes back partly to changes in document collection content reflected by different index term occurrence rates (a.k.a. term frequency), partly to changes in the proportion of referential meaning added to H by sense

definitions and sense relations of index terms. Both T and V can be analyzed by comparing consecutive spectral decomposition of the same index term over periods.

It is key to the understanding of V to remember that the semantic interpretation of both A and AA^T goes back to term occurrences in context, and thereby to the distributional hypothesis of word meaning [30]. However, taking a broader view of the issue, it is clear that at least one more factor, i.e. referential meaning must play a role in interpreting the above matrices as well. Namely the reason *why* terms in a particular context co-occur goes back to their ontological meaning, in a referential relation with their occurrences in sentences. This external, hidden contribution can be measured e.g. by the inverse relationship between the number of intensions (features) of a word vs. its extensions (cardinality of the set of its examples) [31].

Next we bring arguments for the regional nature of semantic content. We will focus on Euclidean space, i.e. the vector space IR model (VSM) and its offsprings, where a natural concern is to ask, is semantic content exactly or inexactly located, i.e. regional? The answer depends on what one wants to model and by which mathematical objects; however, this ambiguity, the potentially dual nature of semantic content will be key to our conclusions.

4 Duality in semantic representation

4.1 The regionality of semantic content – Wave-like representation

In linguistics, the regional nature of word semantics can be best observed on the overlap between word senses displayed as semantic fields [32, 33]. Priss and Old model the underlying, language-independent conceptual regions by neighbourhood lattices [34]. Further the very concern itself is not new, IR and TC having assumed for a long time that the immediate neighbourhood [4] of relevant terms and documents contains related, and therefore important, information, which can be used for e.g. relevance feedback [35]. In a QI setting, Bruza and Woods ascribed word sense disambiguation to the collapse of meaning superpositions due to disambiguating local term context [36]. Further there is an argument in [37] about support vector machines (SVM) linked to quantum disjunctions, the link being regions, meant to solve the problem to be able to say that apple is a kind of fruit (apple is part of the fruit region, i.e. its hyponym), as opposed to modeling that apple and fruit have something

to do with each other. SVMs do this by finding the separating hyperplane, but more research is needed to understand whether the separating hyperplane defines a region. Finally, instead of regions, [38] measure the distance between subspaces spanned by documents by projecting them into one another.

In vector models using position vectors content is usually exactly located. However, reinforcing Dyvik's and Priss and Old's argument, e.g. Erk also argues for the regionality of word meaning, i.e. its inexact location [4]. She departs from the fact that many models of categorization in psychology represent a concept as a region, characterized by feature vectors with dimension weights, and offers two computational models, both of which can host soft region boundaries. Using so-called type vectors as central vectors, each type vector comes with a vector β which defines the importance of each dimension, thus the type vector and its weight vector define a region. Here, regionality implies gradually decreasing similarity between document, query and term vectors. Another model, of sentence formation and called lexical attraction [15], deals with the likelihood of a syntactic relation decaying over distance like a force.

Regionality also manifests itself if term vectors are embedded into an L_2 space, assigning sums of sinusoids or wavelets to each term in the function space [39, 40]. In these models the length of the period or the length of the support controls the inexactness of semantic content, and given that terms are arranged according to a semantic order, this representation may lead to improvement in classification performance.

Hence both terms and sentence components can be considered as having a regional interpretation or aspect as well. Interference (pattern) models using functions to represent semantic content implement this regionality expectation [18, 7] – as do Erk's token vectors for monosemous vs. polysemous words. By these means, the question has to remain undecided. However, the applicability of the concepts of energy, the Hamiltonian, and regional, "smeared out" content beg for a thought experiment which compares exactly located content to particles, inexactly located content to waves, and invokes the parallel with particle-wave duality in QM.

4.2   Meaning and mass – Particle-like representation

The attempt presented here goes back to CM with the implied argument that if QI methods work on language, terms must have "energies" and, in turn, related "mass" equivalents; but then, these must be observable in the much simpler Newtonian environment as well.

Following this train of thought, given that in CM, force is the product of mass and acceleration, $F = ma$ in Newton's 2nd law, we assume that similarity is a "glue", i.e. binding force, between entities in vector space [41]: the more similar two vectors are, the better they attract each other (which is different for example from electromagnetic attraction and repulsion).

Thinking back of Salton's dynamic library with its moving cluster centroids due to collection update (expansion) [42], we leave the question undecided here if an updated system is a closed or open one. With a conservative system in mind, we consider a vector space of many term-document matrices, representing consecutive updates of the same database with a fixed vocabulary over an increasing number of documents, so that all the documents and the terms will have temporal indices as well. This way term and document similarity can be computed as the cosine of two vectors with the same temporal index, and the dislocation (i.e. distance, x) of the same term due to database update will be the cosine of the two respective vectors with consecutive temporal indices. Assuming updates over units of time, term velocity (v) will be the same as the distance between two consecutive positions of the same term. Calculating term acceleration (a) equals the difference between term velocities over units of time. Table 1 shows a toy example of a growing document collection indexed by the same terms in three timesteps. Given this, we depart from Ehrenfest's theorem stating that the time-dependent expectation value of the position operator

$$\langle \psi(t)|x|\psi(t)\rangle = \int \psi(x,t)^* x \psi(x,t) dx$$

satisfies the equation

$$\frac{d^2\langle x(t)\rangle}{dt^2} = \langle \psi(t)|\nabla V(x)|\psi(t)\rangle.$$

This means that the expectation values of the position operator obey their corresponding classical equations of motion, that is, the centre of each wave packet rolls towards the nearest minimum of the potential according to Newton's 2nd law. Following this train of thought further, with term similarities at different times as attraction between them in a symmetric matrix, we insert term acceleration in the respective row and column headings and see that $|F|/|a| = m$ for every term at a given time, i.e. for every term pair compute their time-dependent masses.

We derive term "mass" the following way. Assuming unit time steps, we calculate term velocities between subsequent time steps based on the

$t = 0$	Doping	Football	Performance	Skiing	Training
d_1	5	2	0	0	0
d_2	4	0	0	3	1
d_3	0	0	4	0	5
d_4	6	0	2	0	0
d_5	0	3	0	0	4
$t = 1$					
d_1	5	2	0	0	0
d_2	4	0	0	3	1
d_3	0	0	4	0	5
d_4	6	0	2	0	0
d_5	0	3	0	0	4
d_6	2	3	0	1	1
d_7	1	0	0	4	5
$t = 2$					
d_1	5	2	0	0	0
d_2	4	0	0	3	1
d_3	0	0	4	0	5
d_4	6	0	2	0	0
d_5	0	3	0	0	4
d_6	2	3	0	1	1
d_7	1	0	0	4	5
d_8	5	6	1	1	0
d_9	2	1	1	3	0

Table 1. Evolution of an indexing vocabulary over time

Euclidean distance between the term's vector at the respective time steps. We only care about the magnitude. The change in the velocity results in an acceleration value. At every time step, we regard the force that acts on a term as a sum of dissimilarities with every term except itself. This calculation is similar to an interaction potential that considers $n : n$ relations between particles (or agents) [1, p.157]. This is a considerable simplification as the force that changes the term distribution is at least partially external to an existing distributional pattern. Based on the absolute value of the acceleration and the force, term "mass" can be calculated (see Table 2).

	Doping	Football	Performance	Skiing	Training
v_1	9	9	0	25	36
v_2	49	49	4	16	0
a	40	40	4	-9	-36
F	1.56	1.28	1.24	1.35	1.37
m	0.039	0.032	0.31	0.15	0.038

Table 2. Calculation of term mass over t_0-t_1

5 Discussion

Immediate consequences of term representation by wave packets include the following:

– When describing the behaviour of index terms in a document collection expanding over time according to the CM frame of thought, all the numerical results for variables like force, acceleration or mass, are dimensionless, i.e. metaphoric. An explanation of the results can be to ascribe constant mass to terms like in [41] where term occurrences count as such, but variable density, $\rho = mV$. Here any word form would count as a unit container, hence $V = $ const., and variable term density is the result of variable forces acting on the system in different periods;

– In DQC, Gaussian wave packets model particles [28]. According to Eq 6.4, "the generalized Gaussian packet shows that the packet center moves along the classical trajectory for a particle starting with the given initial mean position and mean momentum". A wave packet (or wave train) is a short "burst" or "envelope" of wave action that

travels as a unit and can be analyzed into, or can be synthesized from, an infinite set of component sinusoidal waves of different wavenumbers, with phases and amplitudes such that they interfere constructively only over a small region of space, and destructively elsewhere. Wave packets, while behaving as particles, can reversibly model sentences as word sequences. If a sentence happens to be a definition of a word sense, like in Wordnet, they model referential semantics charging words "from the outside", i.e. adding extra meaning to words in the term-document matrix whose only meaning this far was ascribed to the distributional hypothesis [30];

– The total energy of a classical mechanical or a QM system is described by the Hamiltonian operator, adding up its kinetic and potential energy. However, the Hamiltonian may or may not describe a quantum-like system. Further there are some subtle issues here and we must not jump to conclusions. Inexactly located semantic content suggests a distribution as well as some kind of an uncertainty; in a QL context both stem from the non-commutative nature of the operators. We only conjecture that the use of such non-commutative operators is an apt description for semantic content.

6   Conclusions

To further underpin the observation that language may be a quantum-like system, we sampled arguments in favour of a metaphorical use of classical and quantum mechanics to model the static and dynamic behaviour of word meaning. With energy as a concept frequently used in some form in text categorization and information retrieval, we contrasted two equally valid approaches to the representation of semantic content, by its exact vs. inexact location, and argued that they closely resemble particle-wave duality in QM. One cited example, wave packets, interestingly behave like particles, lending credibility to the duality hypothesis. Further we showed on a toy example in a classical mechanics framework how metaphorical term mass can be computed. Our future work will increasingly focus on the interpretation and practical use of the Hamiltonian of evolving semantic systems.

References

1. Khrennikov, A.: Ubiquitous quantum structure: from psychology to finance. Springer Verlag (2010)

2. Wittek, P., Darányi, S.: Spectral composition of semantic spaces. In: Proceedings of QI-11, 5th International Quantum Interaction Symposium, Aberdeen, UK (June 2011)

3. LeCun, Y., Chopra, S., Hadsell, R.: A tutorial on energy-based learning. In: Predicting Structured Data. (2006) 1–59

4. Erk, K.: Representing words as regions in vector space. In: Proceedings of CoNLL-09, 13th Conference on Computational Natural Language Learning, Boulder, CO, USA (June 2009) 57–65

5. Park, L., Palaniswami, M., Kotagiri, R.: Internet document filtering using Fourier domain scoring. Lecture notes in computer science (2001) 362–373

6. Park, L.: Spectral Based Information Retrieval. PhD thesis, University of Melbourne (December 2003)

7. Dorrer, C., Londero, P., Anderson, M., Wallentowitz, S., Walmsley, I.: Computing with interference: all-optical single-query 50-element database search. In: Proceedings of QELS-01, Quantum Electronics and Laser Science Conference. (2001) 149–150

8. Buriol, L., Castillo, C., Donato, D., Leonardi, S., Millozzi, S.: Temporal analysis of the wikigraph. In: Proceedings of WI-06, 5th International Conference on Web Intelligence, Hong Kong (December 2006) 45–51

9. Lin, Y., Sundaram, H., Chi, Y., Tatemura, J., Tseng, B.: Detecting splogs via temporal dynamics using self-similarity analysis. ACM Transactions on the Web **2**(1) (2008)

10. Fujimura, K., Fujimura, S., Matsubayashi, T., Yamada, T., Okuda, H.: Topigraphy: visualization for large-scale tag clouds. In: Proceedings of WWW-08, 17th International Conference on World Wide Web, Beijing, China (April 2008) 1087–1088

11. Jurgens, D., Stevens, K.: Event detection in blogs using temporal random indexing. In: Proceedings of the Workshop on Events in Emerging Text Types, Wolverhampton, UK (September 2009) 9–16

12. White, H.: Cross-textual cohesion and coherence. In: Proceedings of the Workshop on Discourse Architectures: The Design and Analysis of Computer-Mediated Conversation, Minneapolis, MN, USA (April 2002)

13. Foltz, P.W., Kintsch, W., Landauer, T.K.: The measurement of textual coherence with latent semantic analysis. Discourse Processes **25**(2-3) (1998) 285–307

14. Gärdenfors, P.: Conceptual spaces: The geometry of thought. The MIT Press (2000)

15. Beeferman, D., Berger, A., Lafferty, J.: A model of lexical attraction and repulsion. In: Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics, Madrid, Spain (July 1997) 373–380

16. Yuret, D.: Discovery of linguistic relations using lexical attraction. arXiv preprint cmp-lg/9805009 (1998)

17. Hutchens, J.: Infomagnetism and sentence generation. Technical report, Artificial Intelligence (2001)

18. Azzopardi, L.: Wave motion: A new metaphor for 2D information visualization the exploration of a metaphor. Technical report, University of Newcastle (2000)

19. Bruce, E.: Biomedical signal processing and signal modeling. Wiley-Interscience (2001)

20. Wang, C., Wang, X.: Indexing very high-dimensional sparse and quasi-sparse vectors for similarity searches. The VLDB journal **9**(4) (2001) 344–361

21. Chatterjee, A., Bhowmick, S., Raghavan, P.: FAST: Force-directed approximate subspace transformation to improve unsupervised document classification. In: Proceedings of 6th Text Mining Workshop held in conjunction with SIAM International Conference on Data Mining, Atlanta, GA, USA (April 2008)
22. Giraud, C.: Gravitational clustering and additive coalescence. Stochastic processes and their applications **115**(8) (2005) 1302–1322
23. Oyang, Y., Chen, C., Yang, T.: A study on the hierarchical data clustering algorithm based on gravity theory. Proceedings of PKDD-01, 5th European Conference on Principles of Data Mining and Knowledge Discovery (September 2001) 350–361
24. Blatt, M., Wiseman, S., Domany, E.: Superparamagnetic clustering of data. Physical Review Letters **76**(18) (1996) 3251–3254
25. Xiao-dan, W., Dian-min, Y., Feng-li, L., Chao-hsien, C.: Distributed model based sampling technique for privacy preserving clustering. In: Proceedings of ICMSE-07, International on Conference on Management Science and Engineering. (August 2007) 192–197
26. Sheikholeslami, G., Chatterjee, S., Zhang, A.: WaveCluster: a wavelet-based clustering approach for spatial data in very large databases. The VLDB Journal **8**(3) (2000) 289–304
27. Horn, D., Gottlieb, A.: The method of quantum clustering. In Dietterich, T., Becker, S., Ghahramani, Z., eds.: Advances in Neural Information Processing Systems. Volume 14., Vancouver, Canada (2001) 769–776
28. Weinstein, M., Horn, D.: Dynamic quantum clustering: A method for visual exploration of structures in data. Physical Review E **80**(6) (2009) 066117
29. Di Buccio, E., Di Nunzio, G.: Envisioning dynamic quantum clustering in information retrieval. In: Proceedings of QI-11, 5th International Quantum Interaction Symposium, Aberdeen, UK (June 2011) 211–216
30. Harris, Z.: Distributional structure. In Harris, Z., ed.: Papers in structural and transformational Linguistics. Formal Linguistics. Humanities Press, New York, NY, USA (1970) 775–794
31. Carnap, R.: Meaning and Necessity: A Study in Semantics and Modal Logic. University Of Chicago Press, Chicago, IL, USA (1947)
32. Lehrer, A.: Semantic fields and lexical structure. American Elsevier, New York, NY, US (1975)
33. Dyvik, H.: Translations as a semantic knowledge source. In: Proceedings of HLT-05, 2nd Baltic Conference on Human Language Technologies, Tallinn, Estonia (April 2005) 27–38
34. Priss, U., Old, L.: Conceptual exploration of semantic mirrors. In: Proceedings of ICFCA-05, 3rd International Conference on Formal Concept Analysis, Lens, France (February 2005)
35. Rocchio, J.: Relevance feedback in information retrieval. In Salton, G., ed.: The SMART Retrieval System: Experiments in Automatic Document Processing. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1971) 313–323
36. Bruza, P., Woods, J.: Quantum collapse in semantic space: interpreting natural language argumentation. In: Proceedings of QI-08, 2nd International Symposium on Quantum Interaction, Oxford, UK (March 2008)
37. Widdows, D.: Semantic vector products: Some initial investigations. In: Proceedings of QI-08, 2nd International Symposium on Quantum Interaction, Oxford, UK (March 2008)
38. Zuccon, G., Azzopardi, L., Rijsbergen, C.: Semantic spaces: Measuring the distance between different subspaces. In: Proceedings of QI-09, 3rd International Symposium on Quantum Interaction, Saarbruecken, Germany (March 2009) 225–236

39. Wittek, P., Darányi, S.: Representing word semantics for IR by continuous functions. In Dominich, S., Kiss, F., eds.: Studies in Theory of Information Retrieval. Proceedings of ICTIR-07, 1st International Conference of the Theory of Information Retrieval, Budapest, Hungary (October 2007) 149–155

40. Wittek, P., Tan, C.L.: Compactly supported basis functions as support vector kernels for classification. Transactions on Pattern Analysis and Machine Intelligence **33**(10) (2011) 2039 –2050

41. Peng, L., Chen, Y., Yang, B., Chen, Z.: A novel classification method based on data gravitation. In: Proceedings of ICNNB-05, International Conference on Neural Networks and Brain, Beijing, China (October 2005) 667–672

42. Salton, G.: Dynamic information and library processing. (1975)