Some steps towards the generation of diachronic WordNets

Yuri Bizzoni
Saarland University
Saarbrücken, Germany
yuri.bizzoni@uni-saarland.de

Marius Mosbach
Saarland University
Saarbrücken, Germany
mmosbach@lsv.uni-saarland.de

Dietrich Klakow
Saarland University
Saarbrücken, Germany
dietrich.klakow@lsv.uni-saarland.de

Stefania Degaetano-Ortlieb
Saarland University
Saarbrücken, Germany
s.degaetano@mx.uni-saarland.de

Abstract

We apply hyperbolic embeddings to trace the dynamics of change of conceptual-semantic relationships in a large diachronic scientific corpus (200 years). Our focus is on emerging scientific fields and the increasingly specialized terminology establishing around them. Reproducing high-quality hierarchical structures such as WordNet on a diachronic scale is a very difficult task. Hyperbolic embeddings can map partial graphs into low dimensional, continuous hierarchical spaces, making more explicit the latent structure of the input. We show that starting from simple lists of word pairs (rather than a list of entities with directional links) it is possible to build diachronic hierarchical semantic spaces which allow us to model a process towards specialization for selected scientific fields.

1 Introduction

Knowledge of how conceptual structures change over time and how the hierarchical relations among their components evolve is key to the comprehension of language evolution. Recently, the distributional modelling of relationships between concepts has allowed the community to move a bit further in understanding the true mechanisms of semantic organization (Baroni and Lenci, 2010; Kochmar and Briscoe, 2014; Marelli and Baroni, 2015), as well as in better mapping language change in terms of shifts in continuous semantic values (Hamilton et al., 2016; Hellrich and Hahn, 2017; Stewart and Eisenstein, 2017). In the past decades, extensive work has also gone into creating databases of hierarchical conceptual-semantic relationships, the most famous of these ontologies probably being WordNet (Miller, 1995). These hand-made resources are tools of high quality and precision, but they are difficult to reproduce on a diachronic scale (Bizzoni et al., 2014), due to word form changes (De Melo, 2014) and shifts in meaning (Depuydt, 2016), which always make it hard to determine "when", over a period of time, a new lexical hierarchy is in place (Kafe, 2017).

A recent attempt to integrate hierarchical structures, typical of lexical ontologies, and the commutative nature of semantic spaces are hyperbolic embeddings (Nickel and Kiela, 2017). Hyperbolic embeddings have shown to be able to learn hierarchically structured, continuous, and low-dimensional semantic spaces from ordered lists of words: it is easy to see how such technology can be of interest for the construction of diachronic dynamic ontologies. In contrast to hand-made resources, they can be built quickly from historical corpora, while retaining a hierarchical structure absent in traditional semantic spaces. In their work Nickel and Kiela (2017) have extensively evaluated hyperbolic embeddings on various tasks (taxonomies, link prediction in networks, lexical entailment), evaluating in particular the ability of these embeddings to infer hierarchical relationships without supervision.

This paper is a first attempt in the direction of using hyperbolic semantic spaces to generate *diachronic lexical ontologies*. While count-based and neural word embeddings have often been applied to historical data sets (Jatowt and Duh, 2014; Kutuzov et al., 2018), and the temporal dimension has even solicited innovative kinds of distributional spaces (Dubossarsky et al., 2015; Bamler and Mandt, 2017), this is to the best of our knowledge the first attempt to model a diachronic corpus through hierarchical, non-euclidean semantic spaces. The literature on hyperbolic embeddings has until now mainly focused on reproducing lexical and social networks from contemporary data (Chamberlain et al., 2017; Nickel and Kiela,

2018).

We demonstrate that these kinds of word embeddings, while far from perfect, can capture relevant changes in large scale lexico-semantic relations. These relations are on the "vertical" axis, defining a super-subordinate structure latent in the data. But we also show that meaningful relations between words are preserved on the "horizontal" axis (similarity of meaning, common semantic belonging) as typically captured by distributional spaces and topic models.

While distributional semantic spaces can be built from unconstrained texts, the main conceptual limitation of hyperbolic embeddings probably lies in the fact the user always needs to precompose (and so pre-interpret) their input in the form of a list of entities linked by a set of parent–children relations; we thus show a simple system to collect *undirected* relations between entities that require less pre-interpretation of the texts at hand and a broader lexical coverage, giving more value to the information provided by the spaces.

Our main contributions are thus two. First, we apply hyperbolic embeddings to a diachronic setting, for which hand-crafted hierarchical resources are extremely difficult to create. Second, we introduce a system to design training inputs that do not rely on directional lists of related word pairs as in previous works. This is particularly advantageous as the system does not need a pre-interpretation nor a pre-formulation of the data in terms of explicit hierarchy and it allows a wider terminological coverage than the previous systems.

2 Methodology

2.1 Data

As our data set, we use the Royal Society Corpus (RSC; version 4.0; Kermes et al. (2016))[1], containing around 10.000 journal articles of the Transactions and Proceedings of the Royal Society in London (approx. 32 million tokens). The time span covered is from 1665 to 1869 and the corpus is split up into five main periods (*1650*: 1665-1699, *1700*: 1700-1749, *1750*: 1750-1799, *1800*: 1800-1849, *1850*: 1850-1869).

As meta-data annotation, the RSC provides e.g. title, author, year, and journal of publication. Crucial for our investigation is the annotation of scientific disciplines (18 in total), which has been approximated by topic modeling (Blei et al., 2003) using Mallet (Fankhauser et al., 2016). Each document is annotated with primary topic and secondary topic, each with confidence scores. We select two groups: (1) the primary topics Chemistry and Physiology, which are subdivided in two sub-groups (Chemistry I and II and Physiology I and II) and thus might indicate more pronounced specialization tendencies, (2) Botany and Galaxy, both forming only one topic each, and thus possibly reflecting less pronounced specialization tendencies. Table 1 presents a detailed corpus statistics on tokens, lemmas and sentences across decades.

decade	tokens	lemmas	sentences
1660-69	455,259	369,718	10,860
1670-79	831,190	687,285	17,957
1680-89	573,018	466,795	13,230
1690-99	723,389	581,821	17,886
1700-09	780,721	615,770	23,338
1710-19	489,857	383,186	17,510
1720-29	538,145	427,016	12,499
1730-39	599,977	473,164	16,444
1740-49	1,006,093	804,523	26,673
1750-59	1,179,112	919,169	34,162
1760-69	972,672	734,938	27,506
1770-79	1,501,388	1,146,489	41,412
1780-89	1,354,124	1,052,006	37,082
1790-99	1,335,484	1,043,913	36,727
1800-09	1,615,564	1,298,978	45,666
1810-19	1,446,900	1,136,581	42,998
1820-29	1,408,473	1,064,613	43,701
1830-39	2,613,486	2,035,107	81,500
1840-49	2,028,140	1,565,654	70,745
1850-59	4,610,380	3,585,299	146,085
1860-69	5,889,353	4,474,432	202,488
total	31,952,725	24,866,457	966,469

Table 1: Corpus statistics of the RSC per decade.

2.2 Approach

Our approach encompasses (1) extraction of relations from data to serve as training data (edge extraction), (2) modeling hyperbolic embeddings on the obtained data, and (3) testing on selected benchmarks.

Edge extraction. In order to select relevant entities, we used the word clusters of a topic model trained on the whole RSC corpus (Fankhauser et al., 2016; Fischer et al., 2018), which generated circa 50 meaningful clusters, mainly belonging to disciplines (such as Paleontology, Electromagnetism) or objects of interest (such as Solar System or Terrestrial Magnetism).

[1] We obtained the RSC from the CLARIN-D repository at http://hdl.handle.net/21.11119/0000-0001-7E8B-6.

topic label	words in topic
Chemistry	acid baro-selenite acid.-when hydroguretted salifiable diethacetone subphosphate meta-furfurol chlorionic causticity acidt acld pyromeconate chloric acids pyroxylic diethyl acid* acid. iodic
Galaxy	stars star to1 nebulosity milky-way facula rethe constellations nebulae lyrce nebula nebule presidencies pole-star st nebulhe sun-spots stars* nebulosities magnet.-

Table 2: The first 20 words from the Chemistry and the galactic Astronomy topic clusters.

For this study, we selected the topics of Chemistry, Physiology, Botany, and galactic Astronomy. Chemistry and Physiology during the time span covered by our corpus undergo a significant inner systematization, which is mirrored by the fact that they are both represented in to two distinct and cohesive topics in our topic model. Botany and galactic Astronomy also underwent major changes during the covered years, but, despite important systematization efforts, kept a more multi-centered conceptual architecture: as a consequence, they represent less cohesive clusters, with more noise and internal diversity. Since the meaningful clusters drawn from topic modeling were relatively small, we populated them through cosine similarity in euclidean semantic spaces built on the same corpus, so as to attain lists of circa 500 elements, of the kind shown in Table 2. Notwithstanding the predictable amount of noise present in these lists, they keep a relative topical cohesion[2].

Based on this selection of words, for each of the five 50-years periods of the RSC, we extract a list of bigrams, i.e. pairs of words of entities of interest.

While usually the training input to model hyperbolic word embeddings is based on directional lists of related word pairs (e.g. the *Hearst patterns* extracted via rule-based text queries (Roller et al., 2018; Le et al., 2019)), we decided to opt for a more "agnostic" method to create input lists for our model.

We consider two words as related if they occur in the same sentence, and we do *not* express any hierarchical value or direction between the words constituting the input lists: the input can be viewed as an undirected graph[3].

On simple cases, this way of extracting undirected edges appears to work well. As an example, in Figure 1 we show the output space of the Wikipedia article on Maslow's Hierarchy of Needs (a very hierarchical topic). In this case, the keywords were selected manually and the text was simple in its exposition of the theory. According to the hierarchy exposed in the article, human needs are as follows: physiological needs (food, water, shelter, sleep), safety (health, financial, well-being), social needs (family, intimacy, friendships), self-esteem, self-actualization (parenting), transcendence. In the hyper-space resulting from this text, the word *needs* occupies the root of the hierarchy: it is the closest point to the origin of the axes and has, consequently, the smallest norm. The six categories of needs described in the input page directly follow as hyponyms: *physiological*, *safety*, *social*, *self-esteem*, *self-actualization*, *transcendence*. The specific kinds of needs mainly cluster as hyponyms of such categories: for example *water, food, sleep, shelter* are all very close in the space, higher in norm, and located as direct hyponyms of *physiological* (they are closer to *physiological* than to the other categories).

The case we are going to deal with in this paper is much more complex: the lists of terms were selected automatically and the corpus is diachronic, technical in nature, and occasionally noisy.

On our corpus, we obtain through our system of edge extraction lists of variable length, between 500 and 5000 pairs depending on the topic and period. While this approach makes the input noisier and the model potentially more prone to errors, the system requires way less starting assumptions on the nature of the data, guarantees a larger coverage than the previous methods, and re-introduces the principle of unstructured distributional profiling so effective in euclidean semantic spaces.

Poincare hierarchical embeddings. For training hyperbolic semantic spaces, we rely on gensim's implementation of Poincare word embeddings. Here, we apply the Poincare hyperspace semantic model recently described by Nickel and Kiela (2017) on each 50-year period of the RSC corpus. We train each model for 20 epochs, di-

[2]Stop words like adverbs, pronouns, determiners and prepositions are also rare in the lists.

[3]Basically, each word pair is twice in the list: (1) word A related to word B, and (2) word B related to word A.

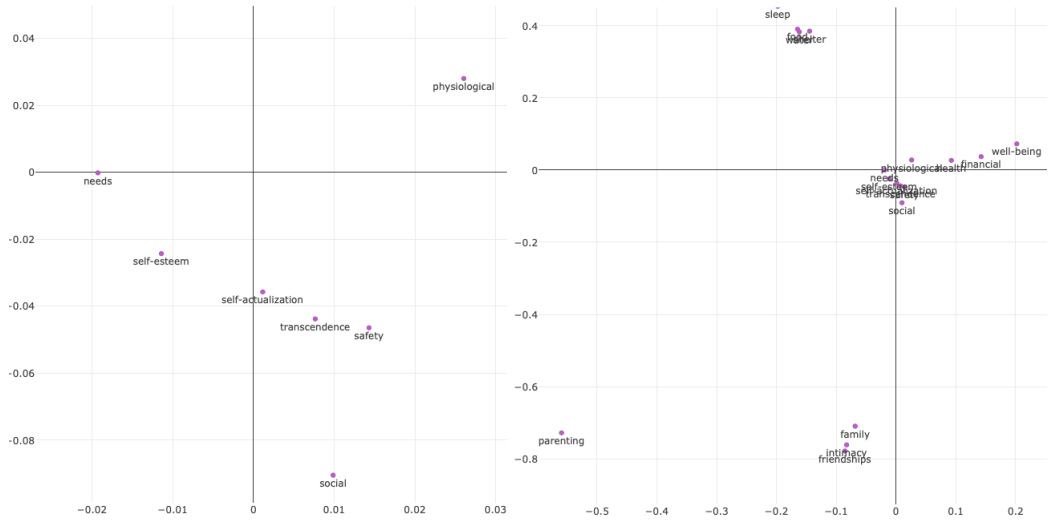

Figure 1: The center of the disk (left) and the whole space (right) as extracted from a Wikipedia article on the Hierarchy of Needs. The main needs cluster around the root of the hierarchy, while their hyponyms cluster to the periphery, but tendentially closer to their hypernymic category than to the others. Note that the space organizes words along the hypernym-hyponym hierarchical line, and ignores other kinds of hierarchy: *physiological*, albeit being treated as more "basic" in the input text, is not closer to *needs* than *transcendence*.

rectly setting a bi-dimensional output. Since our Poincare models generate 2d spaces, we can visualize them without losing any information.

Benchmarks. Since a gold standard to verify the qualities and pitfalls of diachronic hyperbolic semantic spaces is lacking, and it is of not obvious generation, we use two different benchmarks to perform partial tests of the results. The first benchmark is the correlation between the number of WordNet senses and words' norm in the spaces. The other benchmark is the same topic modeling described above: we use it to test whether the words that happen to be in the same topic also cluster together in our spaces.

3 Analysis and results

Having a look at the semantic spaces resulting from the four topics we selected, we can already see that Chemistry and Physiology develop a particularly centralized structure, with few elements in the center and a large crown of peripheral terminology, while Botany and galactic Astronomy return less clear symptoms of their inner ordering.

Figure 2, for example, illustrates hyperbolic embeddings of the Chemistry field for each 50-year period (1650s-1850s). The closer to the center, the more abstract (and potentially ambiguous) the meaning of the words should be, while the more distant from the center, the more we should find specialized terminology. In an ideal semantic hyper-space, the center should represent the real root of the ontology, and its edges should represent the most distant leaves.

In some disciplines (mainly Chemistry and Physiology), we observe the emergence of a clearly centralized and hierarchical evolution, while in others (Biology and Astronomy) we see the development a more multi-central, complicated sort of conceptual organization.

Comparing the evolution of Chemistry with galactic Astronomy (see Figure 3), we can see that the development towards hierarchization does apply to both, but is more pronounced in the Chemistry space.

Figure 4, for clarity, shows only selected labels on the spaces of the 1650s and the 1850s: some words pertaining to the empirical framework, such as *inquires* and *investigations*, and technical terms at various degrees of specificity (still mostly absent in the 1650s space). We see how simple forms of conceptual hierarchization appear in the latter space: for example *compound* moves to the center of the disk, close to a cluster including terms like *substance* and *matter* (and others not included for clarity, such as *solution*), all being more abstract in meaning. *Actions* becomes a hypernym of *investigations* and *inquiries*. Instead, the more spe-

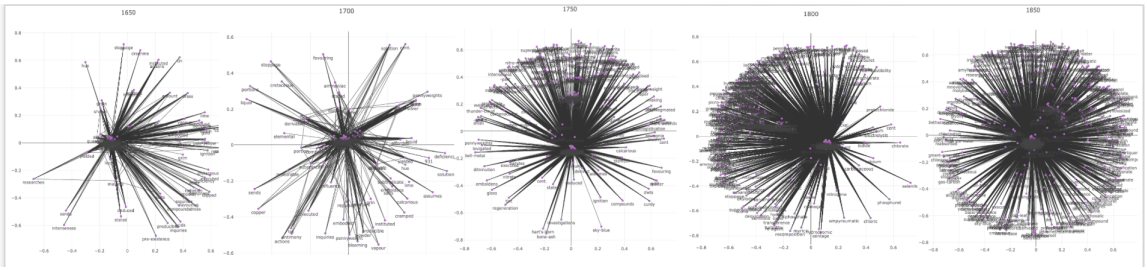

Figure 2: Evolution of the space with original edges for Chemistry.

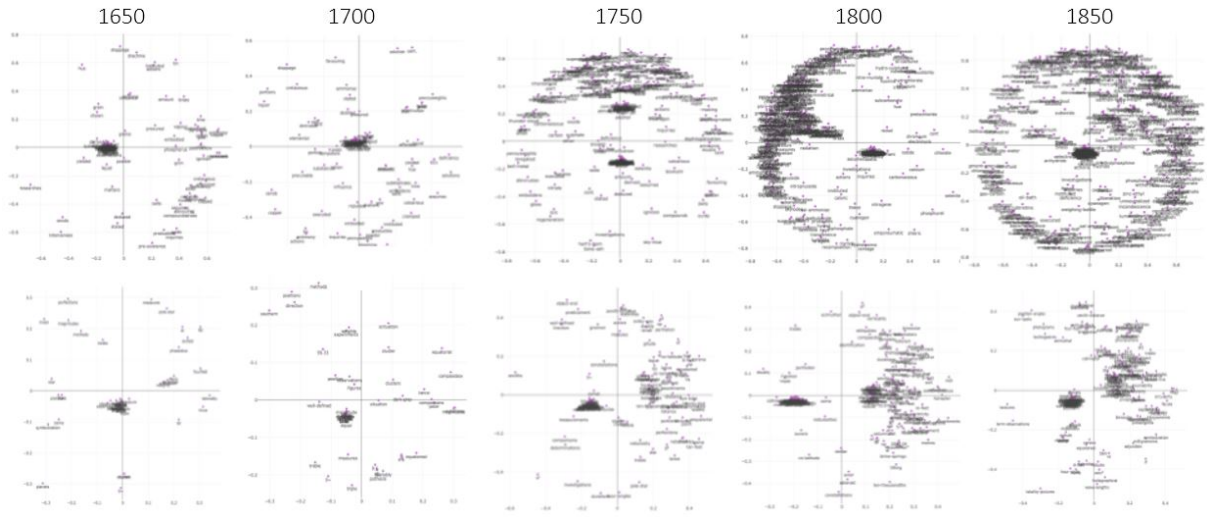

Figure 3: Evolution of the spaces for Chemistry (top row) and galactic Astronomy (bottom row). The high level of hierarchization in Chemistry appears evident. Galactic Astronomy maintains a more chaotic outlook despite the increase of terminology; still, a cluster of terms can be seen growing in the center of the space, while the periphery of the spaces becomes more dense.

cialized terms tend to be located at the edge of the disk, such as *ammoniac* vs. *ammonium-salt*, *anhydride* vs. *carboneous* vs. *gas-carbon*, or *oxide* vs. *protoxide*. See also Table 3 for some examples of developing hierarchization.

This tendency to cluster more clearly abstract/generic and specialized terms is visible in all four disciplines, and is mirrored in the evolution of the structure of the spaces. Measuring the variations in the overall norm of all words, and in the average norm of the 30 elements with the highest and lowest norm of the space for each of the four fields taken into consideration (see Table 4), we record in all cases a tendency to an increasing hierarchization, with small clusters of words moving towards the center and larger numbers of words clustering further away at the periphery of the hyper-disk (see Figure 5 for the highly centralized space of Physiology in the last period of our corpus). Even in Galaxy, the least cohesive of the topics, we notice a steady growth of the average norm (from 3.2 to 20.9), indicating an extension of the periphery. Comparing the results with a "control group" (see again Table 4) formed by sentimental terms (*happiness*, *misery*), which are present throughout the corpus but are neither the topic of the papers nor undergo systematic conceptualizations, there is no hierarchization tendency. Moreover, on average the norm of the 30 most peripheral words steadily increases through time. The tendency of words to increasingly populate more peripheral areas of the disk can be seen as an indication of the increased formation of specialized meanings within particular scientific fields (see Figure 6 for an example).

In Table 4, we show a compendium of these observations for each topic, while in Figure 7 we show the average norm of all words in the space for each discipline through time. It can be seen that the control group does not show most of the trends pictured by the other topics – centralization of a group of words, average increase of the norm,

Epoch	cluster	plant	flowers
1650	clusters, triple, larger	juice, stem, plants	bud, roots, tree
1700	dark-grey, situation, clusters	species, seed, juice	leaves, tree, trees
1750	clusters, nebula, nebulae	flowers, fruit, piece	fruit, branches, plant
1800	nebulosity, clusters, nebulae	leaf-stalk, leaves, roots	shurbs, stem, horse-chestnut
1850	clusters, stellar, nebulae	flowered, seeded, soil	petals, stamina, pistilla

Table 3: Nearest descendants for *cluster*, *plant* and *flowers* in diachronic Poincare spaces for galactic Astronomy (in the first case) and Botany (second and third case). It is possible to observe the emergence of *stellar* as a kind of cluster; of the division between flowered and seeded plants (an antithesis that became meaningful towards the XIX century); and of specific elements of a flower's anatomy, such as the stamen, which were particularly relevant in the studies on flowers' sexuality (mid XIX century).

Figure 4: Selected nodes (in violet) from 1850s Chemistry, as compared to the 1650s. In *Compounds* joins *Substances* and *Matter* to the top of the hierarchy, while *Actions* becomes a hypernym of *Inquires* and *Investigations*. Raw chemical hierarchies can be seen forming at the edges of the hyperdisc.

extension of the peripheries – while a slight trend towards the increase of the norm of the most peripheral words can also be observed in this group.

WordNet comparison. Due to the practical and theoretical difficulties of using contemporary WordNet as a benchmark to validate historical ontologies (should we expect an ideal algorithm to return us a close WordNet similarity on historical data?), we do not use WordNet to directly compare the structure of the spaces (as Nickel and Kiela (2017) do for contemporary data sets), but to correlate the number of WordNet senses a word has with respect to its norm in each period. We notice that in all the considered disciplines, the correlation between the number of senses a word has and its vector's norm is not null, and tends to increase over time (see e.g. Table 5 for Physiology). The words at the center of the hyper-disk tend more and more to overlap with highly polysemous words in contemporary English, while the words that cluster at the edges of the disk correlate more and more with highly specialized words in contemporary English (words with one or two senses at most). Table 5 shows the top 30 words with the lowest norm (most abstract in meaning) and the highest norm (most specialized) for Physiology through time. Both groups show a tendency towards fewer senses over time, indicating increased semantic specialization and decreasing polysemy. Also, in all epochs the first group displays on average more senses than the second group. Table 6 presents Pearson correlation between WordNet senses and words' norms per period across topics, showing an increasing correla-

Epoch	Physiology			Chemistry			Botany			Galaxy			Control		
	H	L	%>.3	H	L	%>.3	H	L	%>.3	H	L	%>.3	H	L	%>.3
1650	0.06	0.53	**45.2**	0.09	0.57	**43.7**	0.10	0.21	**4.3**	0.06	0.20	**3.2**	0.13	0.02	**0.0**
1700	0.11	0.47	**32.4**	0.04	0.44	**33.3**	0.09	0.18	**6.2**	0.02	0.30	**5.3**	0.07	0.01	**0.0**
1750	0.08	0.64	**57.6**	0.09	0.65	**61.2**	0.11	0.43	**3.7**	0.05	0.30	**5.2**	0.10	0.06	**0.0**
1800	0.06	0.68	**67.9**	0.12	0.70	**71.2**	0.10	0.36	**18.0**	0.05	0.35	**15.1**	0.13	0.08	**0.1**
1850	0.06	0.62	**64.1**	0.05	0.69	**69.3**	0.10	0.40	**24.7**	0.04	0.47	**20.9**	0.13	0.07	**0.0**

Table 4: Average norm for the 30 elements with the highest (H) and lowest (L) norm and percentage of elements with norm higher than .3 for each period and discipline.

Epoch	WordNet senses	
	abstract	specialized
1650	11.2	3.4
1700	6.6	4.2
1750	10.9	2.2
1800	5.2	1.03
1850	5.2	0.6

Table 5: Average number of WordNet senses for the 30 terms with the lowest norm (column 2) and for the 30 terms with the highest norm (column 3) in the space of Physiology.

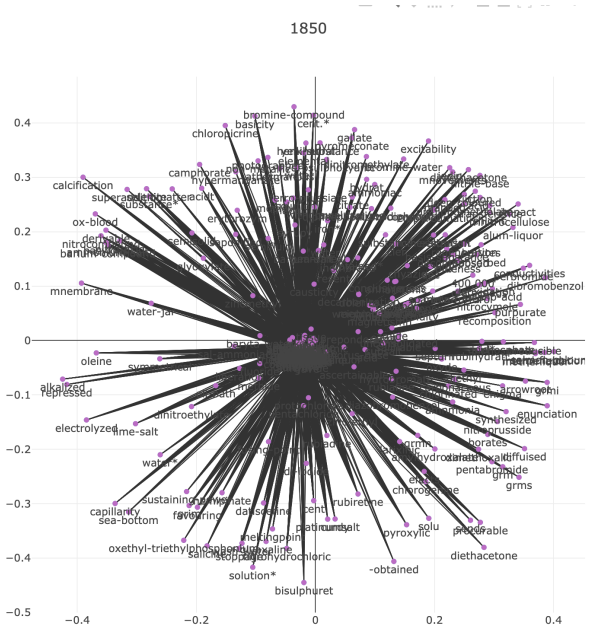

Figure 5: Physiology space (with original edges) for the last period. The centralized hierarchical structure is clearly visible.

tion.

Topic clustering. All four the selected topics show a tendency to increase their words' average norm and the distance between the center and the edge of the disk. The two topics that show stronger symptoms of conceptual hierarchization, Chemistry and Physiology, were also distinguished in two lexical sub-topics by our original topic model. The emergence of these sub-topics was mainly due to the changes in word usage caused by relevant scientific discoveries (like for example the systematization of elements in Chemistry) that created vocabularies and conceptual systems that had scarce interactions with one another. In Table 7, we show that the average cosine similarity between the words belonging to the one sub-topic tends to stay higher than their average similarity to the words belonging to the other sub-topic: the topical distance between the two groups is not lost in the hierarchization.

4 Discussion

We have built diachronic semantic hyperspaces for four scientific topics over a large historical English corpus stretching from 1665 to 1869. We have shown that the resulting spaces present the characters of a growing hierarchization of concepts, both in terms of inner structure and in terms of light comparison with contemporary semantic resources (growing Pearson correlation between norm and WordNet senses). We have shown that while the same trends are visible in all four disciplines, Chemistry and Physiology present more accentuated symptoms of hierarchization, while the group of control had even few or no signs of hierarchization.

Specialization in scientific language. This work is part of a larger project aimed to trace the linguistic development of scientific language toward an optimal code for scientific communication (Degaetano-Ortlieb and Teich, 2018, 2019). One mayor assumption is the diachronic development towards specialization – as a scientific field develops, it will become increasingly specialized and expert-oriented.

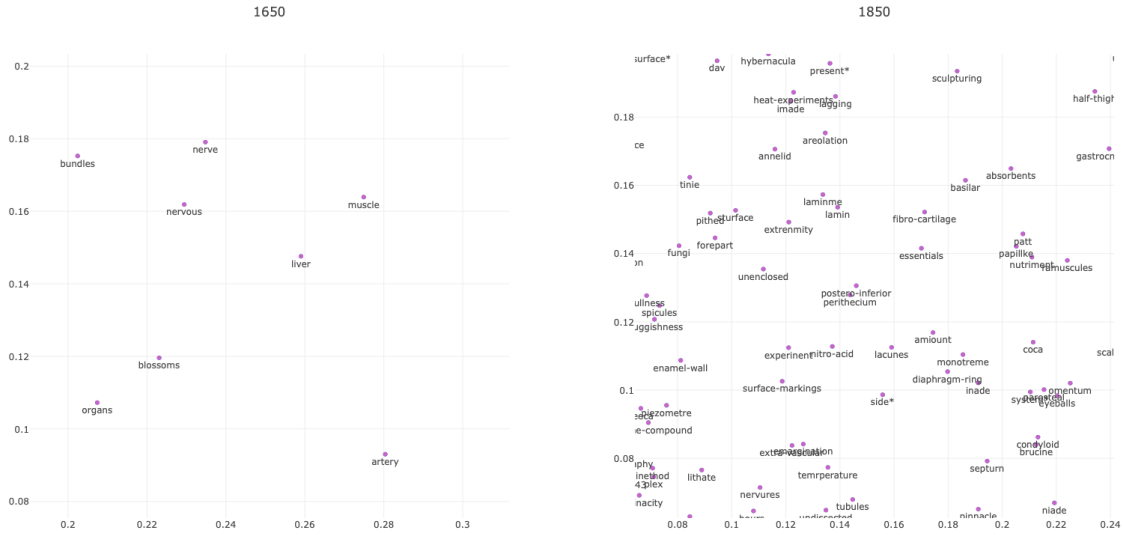

Figure 6: Population of the same area of the hyper-disk for Physiology in the first and last epoch. More specialized and technical terms tend to populate the same level in the "hierarchy".

Epoch	Physiology I and II	Chemistry I and II	Galaxy	Botany	Control
1650	-0.37	-0.42	-0.50	-0.09	-0.06
1700	-0.20	-0.44	-0.35	-0.05	0.67
1750	-0.40	-0.45	-0.43	-0.24	-0.34
1800	-0.42	-0.46	-0.16	-0.22	-0.17
1850	-0.41	-0.46	-0.37	-0.32	-0.16

Table 6: Pearson correlation between WordNet senses and word's norm per period per topic.

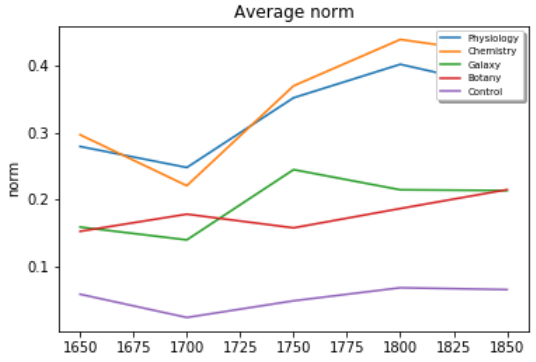

Figure 7: For all four disciplines the average words' norm increases through time. This is due to the expansion of the hyperspace periphery: words become more specialized, hierarchies become deeper. The control group (sentiment terms) does not show this tendency.

Thus, as a field specializes, it develops more technical and differentiated vocabulary (Halliday, 1988; Teich et al., 2016). For the disciplines investigated here increased specialization over time appears clearly in our hyperspaces showing a tendency towards the use of more peripheral words and deeper hierarchies.

Epoch	P in	P out	C in	C out
1650	.58	.59	.54	.55
1700	.60	.60	.56	.56
1750	.53	.53	.50	.49
1800	.51	.50	.48	.47
1850	.50	.47	.47	.44

Table 7: Topic detectability. Average cosine similarity for elements pertaining to the same sub-topic (in) and elements pertaining to different sub-topics (out) in Physiology (P) and Chemistry (C) through time.

Considerations on validity of our baselines. Finding valid, meaningful baselines to evaluate hierarchies based on a diachronic corpus is not a trivial task. Comparing them to the topic

model's results on the same corpus was possibly the most sensible one, but we should not expect too much on that side: Hyperbolic embeddings are not specifically designed to tell topics apart, and if words pertaining to slightly different topics (such as two kinds of chemistry) happen to be on the same level of conceptual abstraction, it is fair to expect them quite near in the hyper-disk geography.

At the same time, comparing our results to WordNet makes sense only partially: the conceptual structures of WordNet are 150 years more recent than the ones discussed in the most recent of our spaces, and it is wrong to assume a priori that their distribution in a historical hierarchy should be similar. So we relied on internal analysis and qualitative considerations, but baselines for these kinds of tasks would be highly needed to better test diachronic ontologies.

Considerations on our extraction system. To collect our data, we used a very simple and non-committal approach that feeds the models with less information than usually provided in the literature.

However, choosing the words with some care and working on large numbers, our models do not seem completely at a loss in front of the noise of the input data. With differences due to the noise of the word lists and the development of the fields, a tendency for specialized terms to cluster as hyponyms of more abstract and polysemous words could be observed in all four disciplines. In future work, we intend to accurately test this procedure by means of contemporary data sets.

Dynamic diachronic WordNets. Hand crafted, historical ontologies of concepts are extremely expensive in terms of person/hour, not considering the amount of expertise and skills required to build a hierarchy of concepts based on the knowledge and beliefs of a different time. We speculate that these sorts of technologies can be a step towards an easier, and more dynamic way of building corpus-induced ontologies, offering for example raw material to be polished by human experts.

References

Robert Bamler and Stephan Mandt. 2017. Dynamic word embeddings. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 380–389. JMLR. org.

Marco Baroni and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *American Journal of Computational Linguistics*, 36(4):673–721.

Yuri Bizzoni, Federico Boschetti, Harry Diakoff, Riccardo Del Gratta, Monica Monachini, and Gregory R Crane. 2014. The making of ancient greek wordnet. In *LREC*, volume 2014, pages 1140–1147.

David M. Blei, Andrew W. Ng, and Michael I. Jordan. 2003. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3:993–1022.

Benjamin Paul Chamberlain, James Clough, and Marc Peter Deisenroth. 2017. Neural embeddings of graphs in hyperbolic space. *arXiv preprint arXiv:1705.10359*.

Gerard De Melo. 2014. Etymological wordnet: Tracing the history of words. In *Proceedings of LREC 2014*, pages 1148–1154.

Stefania Degaetano-Ortlieb and Elke Teich. 2018. Using relative entropy for detection and analysis of periods of diachronic linguistic change. In *Proceedings of the 2nd Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature at COLING2018*, pages 22–33, Santa Fe, NM, USA.

Stefania Degaetano-Ortlieb and Elke Teich. 2019. Toward an optimal code for communication: The case of scientific English. *Corpus Linguistics and Linguistic Theory*, 0(0):1–33. Ahead of print.

Katrien Depuydt. 2016. Diachronic semantic lexicon of dutch (diachroon semantisch lexicon van de nederlandse taal; diamant). In *DH*, pages 777–778.

Haim Dubossarsky, Yulia Tsvetkov, Chris Dyer, and Eitan Grossman. 2015. A bottom up approach to category mapping and meaning change. In *NetWordS*, pages 66–70.

Peter Fankhauser, Jörg Knappen, and Elke Teich. 2016. Topical diversification over time in the royal society corpus. Digital Humanities 2016, Krakow 1116 July 2016, Krakow. Jagiellonian University; Pedagogical University.

Stefan Fischer, Jorg Knappen, and Elke Teich. 2018. Using topic modelling to explore authors' research fields in a corpus of historical scientific english. In *Proceedings of DH 2018*.

M.A.K. Halliday. 1988. On the Language of Physical Science. In Mohsen Ghadessy, editor, *Registers of Written English: Situational Factors and Linguistic Features*, pages 162–177. Pinter, London.

William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Cultural shift or linguistic drift? comparing two computational measures of semantic change. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Conference on Empirical Methods in Natural Language Processing*, volume 2016, page 2116. NIH Public Access.

Johannes Hellrich and Udo Hahn. 2017. Exploring diachronic lexical semantics with JeSemE. In *Proceedings of ACL 2017, System Demonstrations*, pages 31–36, Vancouver, Canada. Association for Computational Linguistics.

Adam Jatowt and Kevin Duh. 2014. A framework for analyzing semantic change of words across time. In *Proceedings of the 14th ACM/IEEE-CS Joint Conference on Digital Libraries*, pages 229–238. IEEE Press.

Eric Kafe. 2017. How stable are wordnet synsets? In *LDK Workshops*, pages 113–124.

Hannah Kermes, Stefania Degaetano-Ortlieb, Ashraf Khamis, Jörg Knappen, and Elke Teich. 2016. The Royal Society Corpus: From Uncharted Data to Corpus. In *Proceedings of the 10th LREC*, Portorož, Slovenia. ELRA.

Ekaterina Kochmar and Ted Briscoe. 2014. Detecting learner errors in the choice of content words using compositional distributional semantics. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 1740–1751.

Andrey Kutuzov, Lilja Øvrelid, Terrence Szymanski, and Erik Velldal. 2018. Diachronic word embeddings and semantic shifts: a survey. *arXiv preprint arXiv:1806.03537*.

Matt Le, Stephen Roller, Laetitia Papaxanthos, Douwe Kiela, and Maximilian Nickel. 2019. Inferring concept hierarchies from text corpora via hyperbolic embeddings. *arXiv preprint arXiv:1902.00913*.

Marco Marelli and Marco Baroni. 2015. Affixation in semantic space: Modeling morpheme meanings with compositional distributional semantics. *Psychological review*, 122(3):485.

George A. Miller. 1995. Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Maximilian Nickel and Douwe Kiela. 2018. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. *arXiv preprint arXiv:1806.03417*.

Maximillian Nickel and Douwe Kiela. 2017. Poincaré embeddings for learning hierarchical representations. In *Advances in neural information processing systems*, pages 6338–6347.

Stephen Roller, Douwe Kiela, and Maximilian Nickel. 2018. Hearst patterns revisited: Automatic hypernym detection from large text corpora. *arXiv preprint arXiv:1806.03191*.

Ian Stewart and Jacob Eisenstein. 2017. Making" fetch" happen: The influence of social and linguistic context on nonstandard word growth and decline. *arXiv preprint arXiv:1709.00345*.

Elke Teich, Stefania Degaetano-Ortlieb, Peter Fankhauser, Hannah Kermes, and Ekaterina Lapshinova-Koltunski. 2016. The Linguistic Construal of Disciplinarity: A Data Mining Approach Using Register Features. *Journal of the Association for Information Science and Technology (JASIST)*, 67(7):1668–1678.