Meaning change in a petri dish: Constructions, semantic vector spaces, and motion charts

Hilpert, Martin; Perek, Florent

DOI:
10.1515/lingvan-2015-0013

Martin Hilpert* and Florent Perek

Meaning change in a petri dish: constructions, semantic vector spaces, and motion charts

DOI 10.1515/lingvan-2015-0013
Received June 15, 2015; accepted September 21, 2015

Abstract: This paper explores how the visualization tool of motion charts can be used for the analysis of meaning change in linguistic constructions. In previous work, linguistic motion charts have been used to represent diachronic frequency trends and changes in the morphosyntactic behavior of linguistic units. The present paper builds on that work, but it shifts the focus to the study of semantic change. How can motion charts be used to visualize semantic change over time? In order to answer this question, we draw on semantic vector space modeling to visualize aspects of linguistic meaning. As an analogy to this approach, the title of this paper alludes to a petri dish in which the growth and development of biological micro-organisms can be observed. On the basis of diachronic corpus data, we monitor developments in the semantic ecology of a construction. This allows us to observe processes such as semantic broadening, semantic narrowing, or semantic shift. We illustrate our approach on the basis of a case study that investigates the diachrony of an English construction that we call the 'many a NOUN' construction.

Keywords: corpus linguistics, semantic change, construction grammar, motion charts, semantic vector spaces

1 Introduction

This paper explores how the visualization tool of motion charts, i.e. 'animated' scatterplots that show temporal developments, can be used for the analysis of meaning change in linguistic constructions. In previous work (Hilpert 2011, 2012a, 2013), motion charts have been used to represent diachronic frequency trends and changes in the relative similarities between sets of linguistic units. An approach of this kind can reveal different kinds of developments, for instance that a verb diachronically gravitates towards one syntactic subcategorization frame, at the expense of other ones that are available (2011: 455). It can also be shown that in a paradigm of constructions, certain members of the paradigm become more and more similar in terms of their morphosyntactic behavior, whereas the remaining members dissimilate (2013: 197). The present paper builds on these studies, but it shifts the focus to the study of semantic change. How can motion charts be used to visualize semantic change over time? In order to answer this question, we draw on an approach developed in Perek (2014, to appear), which exploits the technique of semantic vector space modeling to visualize processes of meaning change. As an analogy to this approach, the title of this paper alludes to a petri dish in which the growth and development of biological microorganisms can be observed. On the basis of diachronic corpus data, we monitor developments in the semantic ecology of a construction. A similar approach has been taken by Sagi et al. (2012), who use Latent Semantic Analysis in order to study processes of lexical semantic change such as semantic broadening (a successively larger area of a semantic space being populated) and semantic narrowing (the populated area is shrinking). As will be discussed in more detail below, our own results are less indicative of semantic narrowing or broadening, but can rather be interpreted in terms of semantic shift (different areas being populated at different points in time). How exactly this kind of analysis works in practice will be illustrated on the basis of a case study that investigates the diachrony of an English construction that we call the *many a NOUN* construction (Hilpert 2012b).

***Corresponding author: Martin Hilpert,** Université de Neuchâtel – English, Espace Louis Agassiz 1, Neuchâtel 2000, Switzerland, E-mail: martin.hilpert@unine.ch
Florent Perek, Universität Basel – Englisches Seminar, Basel, Switzerland

The remainder of this paper is structured as follows. Section 2 presents the *many a NOUN* construction, motivates our interest in the diachrony of that construction, and discusses the diachronic data that forms the basis for our study. Section 3 explains how we constructed a semantic vector space that serves as our 'petri dish', i.e. as the semantic frame of reference against which we assess diachronic shifts in the semantics of the construction. Section 4 discusses our results. To preview the main finding, we observe that the construction loses some of its semantic facets, but that overall, the construction continues to populate a wide semantic area. We argue that this sustained semantic broadness accounts for the fact that the *many a NOUN* construction remains highly productive, even as it wanes in type and token frequency. Section 5 concludes and offers a few pointers for extensions and refinements of our approach.

2 The *many a NOUN* construction

The construction that we will use to illustrate our approach is the *many a NOUN* construction, which is exemplified by phrases such as *many a time* or *many a day*. The construction is mentioned as a possible use of the quantifier *many* in several standard reference works (Visser 1963: §93; Huddleston and Pullum 2002: 394; Quirk et al. 1985: Section 5.23). Comments on it are sparse, but Huddleston and Pullum remark that it is "somewhat formal and archaic" (2002: 394) in Present-Day English. In line with this observation, the construction is the occasional subject of internet forum discussions in which L2 learners of English inquire about meaning and use of the construction.[1] The examples below illustrate the construction with some context.

(1) Katy's face wore a smile of triumph, as Michael was dismissed. Her mother's truthfulness had been vindicated, and it was the proudest moment she had known for **many a day**. (COHA, 1850)

(2) How the men did cheer them! – men who knew what war was by experience; fresh from such fields as Chancellorsville and Gettysburg, and going on to much more and closer work, with few chances in favor of a safe return. It is not strange that **many a father's eye** filled with tears, and **many a rough face** softened into a pleasant smile, as these little ones bade them welcome, and kissed them good-bye. (COHA 1876)

Searches in the British National Corpus and the Corpus of Contemporary American English (Davies 2008) show that the construction is relatively more frequent in written language, and here particularly in fiction and magazine journalism. Interestingly, highly formal genres such as academic writing actually show lower frequencies than the latter. Also, the construction is by no means restricted to written language. For our analysis, we use a dataset from Hilpert (2012b) that is based on the Corpus of Historical American English (Davies 2010). Hilpert exhaustively retrieved word sequences from the COHA in which the quantifier *many* was followed by an indefinite determiner (*a*, *an*), and an element that was tagged as a noun. Up to two intervening elements between the determiner and the noun (as in *many a father's eye*) were allowed to occur. This procedure yielded more than 15,000 hits that represent 3,140 different noun types. For each of these noun types, the token frequencies for each decade of the COHA are represented in the dataset. A small part of the dataset is shown in Table 1 (adapted from Hilpert 2012b: 237).

The full database thus allows us to determine type and token frequencies for all nouns that occur in the construction between 1810 and 2009. A look at the aggregated token frequencies indicates that the construction has undergone a substantial frequency decrease over the past 150 years. That decrease is shown in Figure 1.

As the initial decades of the COHA hold much smaller amounts of text than later ones, and also their genre composition differs from later decades, so that the first few observations should be taken with a grain

1 http://forum.wordreference.com/threads/many-a-time-when.911327/, http://www.usingenglish.com/forum/threads/197476-Many-a-boy-and-many-a-girl, https://www.englishforums.com/English/ManyAManAndManyMen/vndxk/post.htm

Table 1: Noun frequencies in the *many a NOUN* construction from the COHA.

NOUN	1810s	1820s	1830s	1840s	1850s	...
time	2	43	50	78	74	...
day	1	9	52	58	66	...
man	5	14	39	38	61	...
year	3	23	30	54	57	...
...	...	...	...	...	...	...

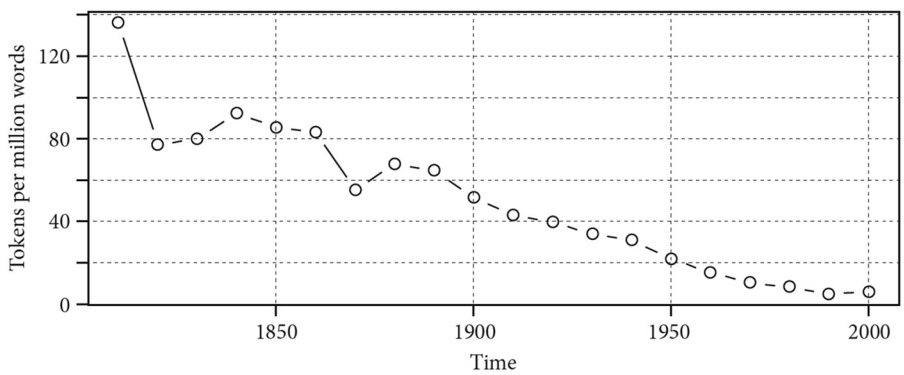


Figure 1: Declining token frequencies of the *many a NOUN* construction.

of salt. However, the overall trend is very clear. Interestingly, the decline in frequency has left the productivity of the construction unaffected. Speakers can be frequently observed to use the construction with new noun types, as in *many a Labrador-owner* or *many a sausage roll* (examples from internet pages). Hilpert (2013: 39) reports that out of 65 types that are observed in the 1990s, no less than 55 are hapax legomena. This contrasts with other receding constructions. For example, the construction that has generated so-called a-adjectives such as *afloat* or *asleep* (Boyd and Goldberg 2011) is no longer available to speakers to form utterances such as *??The car is apark* or *??The children are asit*. Our study thus aims to shed light on the question of why the productivity of *many a NOUN* remains strong. In his study of the construction, Hilpert (2012b) presents a collocational analysis that applies distinctive collexeme analysis (Gries et al. 2004). He finds that different semantic categories dominate in the *many a NOUN* construction at different points in time. In the nineteenth century, the construction displays a strong connection to nouns that evoke human emotions (*heart*, *tear*, *sigh*, etc., cf. *many a father's eye filled with tears* in the example above). During the early twentieth century, time nouns (*day*, *night*, etc.) emerge as the predominant lexical class. Besides those two categories, a semantic class of human beings, illustrated by nouns such as *citizen*, *reader*, or *Republican*, gains in importance during the twentieth century (Hilpert 2012b: 241). In this paper, we revisit these results, but we take a different analytical approach in order to understand better why the *many a NOUN construction* remains productive. In particular, we create a visualization of the shifts that characterize the semantic developments of the construction, and we test whether the observable changes are in line with the conclusions of earlier research.

3 A semantic vector space of the *many a NOUN* construction

In order to study semantic change in the *many a NOUN* construction, we take a distributional approach, specifically semantic vector space modeling (Turney and Pantel 2010). A common linguistic application of semantic vector spaces is the assessment of mutual semantic similarities within a set of lexical elements,

which has been done for instance in Heylen and Ruette (2013), and Ruette et al. (2013). Each lexical element is represented in terms of a collocate frequency vector, that is, a frequency list of those words that are found in close proximity of the element. To give a basic example, a lexical element such as *hotel* would have a collocate vector in which words such as *room*, *grand*, or *elevator* are particularly frequent. These preferences are partially shared with semantically related elements such as *lobby*, but not to the same extent with semantically unrelated words such as *guitar* or *bird*. A semantic vector space of a set of lexical elements can be visualized as a two-dimensional scatterplot that represents a 'semantic landscape', or, to stay with the metaphor of our paper title, a 'petri dish' that represents the meanings that are covered by the set of lexical elements. On such a scatterplot, *hotel* and *lobby* would be placed in relatively close proximity, while *guitar* and *bird* would appear in more distant positions.

The semantic vector space of the *many a NOUN* construction that we construct in this study starts with the nouns that occur in the construction: We are interested in the semantic spectrum of those nouns. We adopt the theoretical perspective of Construction Grammar (Goldberg 1995, 2006; Hilpert 2014), which means that we subscribe to the ideas that syntactic patterns are inherently meaningful (Goldberg 1995: 1) and that the meanings of constructions are closely related to the lexical items that commonly occur within a construction (Goldberg 1995: 50; Stefanowitsch and Gries 2003). We model the semantic spectrum of the *many a NOUN* construction on the basis of synchronic corpus data, specifically the Corpus of Contemporary American English (Davies 2008). The mutual similarities of our noun types is assessed via similarities and differences in their respective collocational preferences in the COCA.[2] In order to construct our semantic vector space, we took the following steps, as adapted from Perek (2014, to appear).

1. We took the 3,140 noun types from our database and selected all types that occurred at least ten times with the construction, which gave us 230 nouns. This restriction was imposed for two reasons. First, having a manageable number of lexical types allows for a manual semantic analysis of those types; second, we did not want our visualizations to become overly cluttered and complex. Our selection of the most frequent elements still retains more than 60% of all datapoints in the database.
2. For each of our 230 nouns, we retrieved exhaustive concordances with four words to the left and four words to the right from the COCA. This step was done in an automated fashion with a full-text version of the COCA that was downloaded to a local computer.
3. From those concordances, we removed punctuation and highly common words, using a list of 150 stopwords that included both function words and a few highly frequent content words.[3] Eliminating stopwords is a common procedure that helps to reduce noise in the statistical analysis, since function words such as *the* and *and* and other frequent items do not correspond to relevant semantic distinctions to the same extent as content words such as *crowd* and *water* (cf. Turney and Pantel 2010: 154).
4. We removed all collocates that occurred less than 100 times with all of the nouns, taken together. This was done to keep the computational effort manageable and also in order to favor collocates that are likely to occur with several of the target nouns.
5. From the cleaned concordances, we put the collocates into a table in which the noun types were the column labels (230) and their collocates were in the rows (5,293). The cells of that table were filled with the observed co-occurrence frequencies of all database items with all collocates. For instance, our database contains the noun *age*, which co-occurs 3 times with the word *Victorian* and 19 times with the word *sixteen* in the COHA when a text window of 4L and 4R is used.

2 Our reliance on synchronic corpus data for this part of the study means that we make a simplifying assumption about our noun types and their respective collocates: As we use the present-day meanings of those elements as a constant semantic frame of reference for the diachronic developments that we observe, we thereby stipulate that our nouns and their collocates have remained semantically stable, assuming that the meaning of the word *year* in the 1810s was the same as it is now. For many words (*woman*, *gay*, *tourism*, etc.), lexical semantic change is clearly in evidence. This limitation has to be kept in mind.

3 The list of stop words is included in our R code, which is available upon request.

6. We took the raw co-occurrence frequencies in the table and weighted them by applying pointwise mutual information (PMI). This is done to control for the differences in text frequency between the database items, and give more weight to collocates that are especially distinctive between words.
7. We set all negative PMI values to zero, so that only positive values remain in our table. The use of positive PMI (PPMI) has been shown to yield superior results in studies of lexical semantic similarity (Bullinaria and Levy 2007).
8. Finally, we created a visualization of the resulting semantic vector space. A popular choice for such visualizations is metric multidimensional scaling (MDS), also used by Perek (2014, to appear). However, MDS shows limitations when dealing with high-dimensional datasets, and is therefore not the best option for distributional semantic spaces in many cases. In what follows, we use t-distributed Stochastic Neighbor Embedding (or t-SNE), which is a dimensionality reduction algorithm that is widely used in distributional semantics (cf. Van der Maaten and Hinton 2008). The t-SNE algorithm has the property of preserving the local similarity relations of the original dataset, and is therefore very good at identifying clusters of highly similar items.

The resulting two-dimensional scatterplot of noun types is shown in Figure 2.[4] The graph places elements with strong semantic relations in close mutual proximity and places semantically unrelated elements at a greater mutual distance. For now, we disregard the frequencies of the noun types, which will be examined

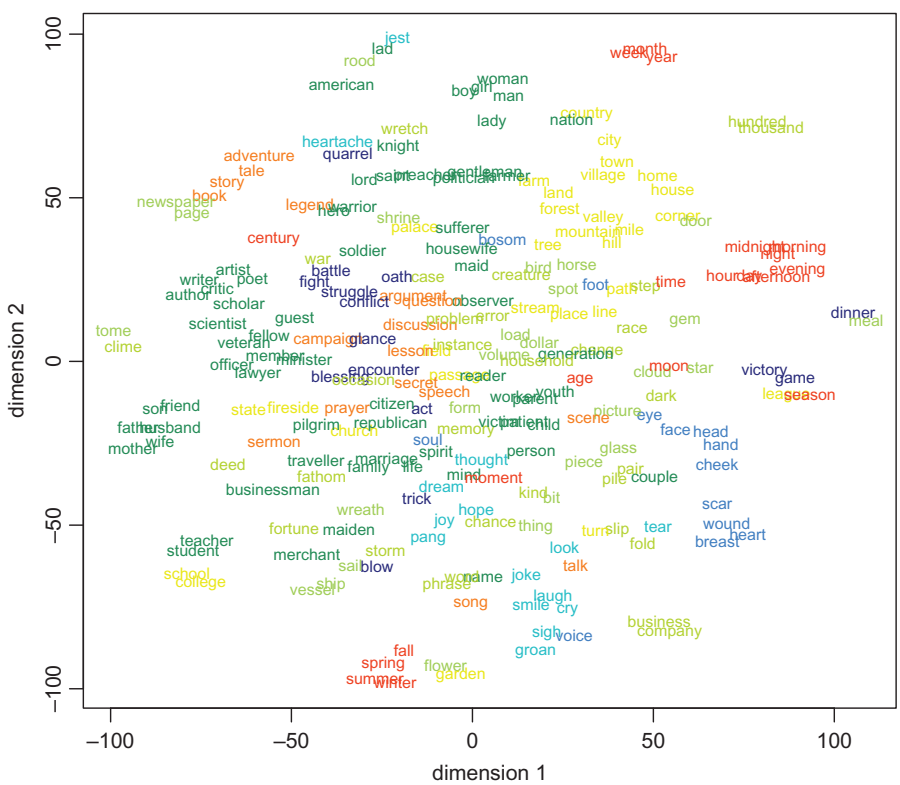


Figure 2: Semantic vector space of *many a NOUN* based on data from the COCA.

4 Figure 2 was generated by means of the tsne R package, using 5,000 iterations of the t-SNE algorithm with a perplexity parameter of 50.

further below, in order to focus on semantic aspects. At its most basic, the graph reflects the closeness in meaning of words such as the body part terms *eye*, *face*, *head*, *hand*, *cheek*, and others (shown in light blue, x = 75, y = −25), family-related terms such as *father*, *wife*, *husband*, *son*, and *friend* (shown in dark green, x = −90, y = −25), the four seasons *spring*, *summer*, *fall* and *winter* (in red, x = −25, y = −100), and times of the day such as *afternoon*, *morning*, *night*, *midnight*, and *evening* (also in red, x = 80, y = 25).

The color-coding of the nouns in Figure 2 represents a categorization into semantic classes. In studies that use semantic vector spaces, it is a common analytical step to group the elements under analysis into different semantic categories, which can be done using inductive methods such as hierarchical clustering. For example, Perek (2014, to appear) clusters the verbs of an English argument structure construction over their collocational behaviors, arriving at a four-fold categorization that distinguishes psych-verbs (*surprise*), abstract actions (*explain*), physical actions (*squeeze*), and a remainder category. In the present study, we opt for a manual, intuitive classification into the nine categories that are shown in Table 2. Our motivation for integrating human judgments into our analysis is that we want to see how the t-SNE algorithm compares to our own intuitions.

Table 2: Semantic categories used in the present analysis.

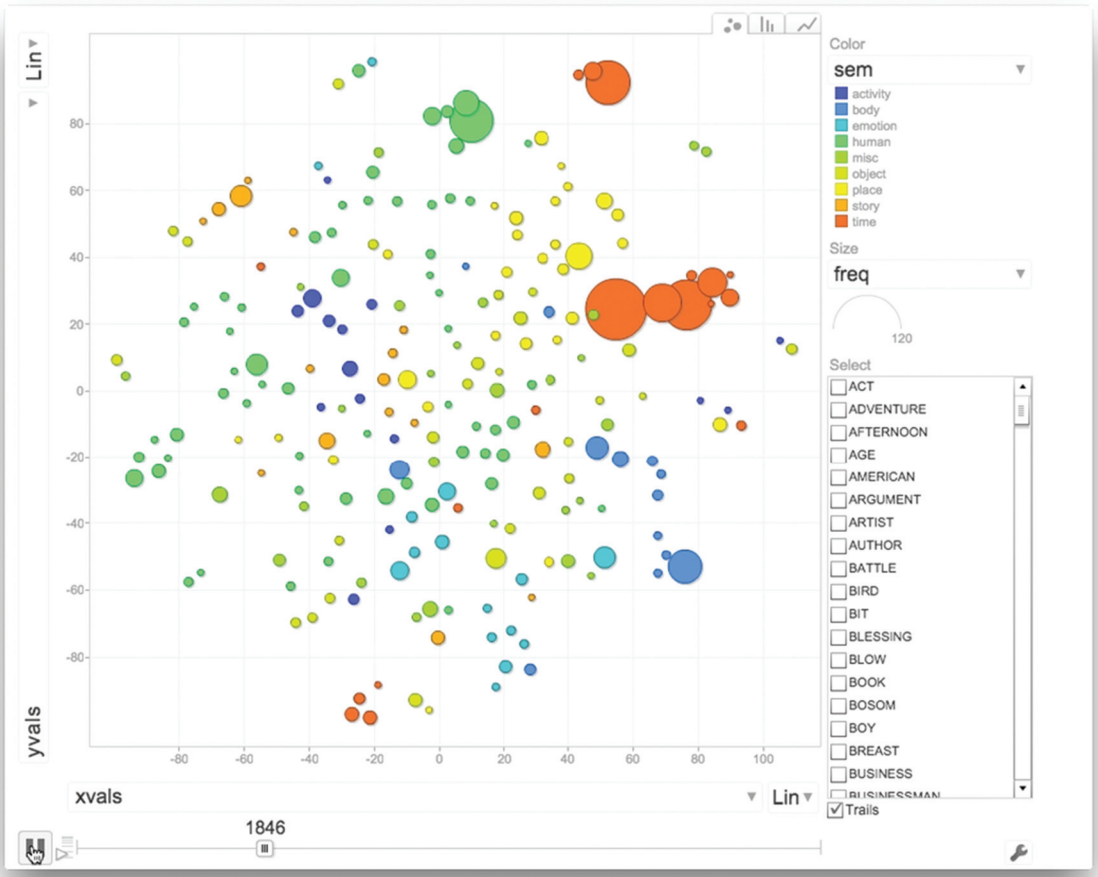
semantic category	examples	color
time	time, day, afternoon, winter	red
human being	man, wife, friend, lawyer	dark green
body part	eye, face, foot, head	light blue
emotion-related	cry, groan, hope, sigh	turqoise
place	town, valley, village, farm	yellow
object	ship, flower, door	olive
story	story, tale, legend, book	orange
activity	battle, dinner, game	purple
miscellaneous	company, error, war, problem	lime

Figure 2 makes evident that our manual semantic categorization of the 230 nouns does not lead to a cleanly distinguished set of clusters with identical colors. Nonetheless, the t-SNE algorithm shows a remarkable sensitivity to semantic detail: For instance, our manual categorization groups together seasonal terms (*spring*, *summer*) with time-of-day expressions (*afternoon*, *night*). The t-SNE algorithm reveals that in terms of their collocates, these nouns actually do not belong together. Similarly, Figure 2 shows that the noun *season*, which we also categorized as a time expression, is more strongly related to the nouns *game*, *victory*, and *league*, due to its sense as a 'playing season' in sports. The graph further shows, nonsurprisingly, that the categories of objects (orange) and activities (yellow) are not restricted to a small part of the semantic vector space. Rather, these nouns are scattered over the entire area of the graph. These issues notwithstanding, Figure 2 contains a lot of semantic structure, which makes it useful as a basis for further investigations of the *many a NOUN* construction. In the next section, we discuss diachronic developments in the nine semantic categories that become apparent when the semantic vector space is projected onto the diachronic frequency data that we have collected from the COHA.

4 The *many a NOUN* construction over time

How has the *many a NOUN* construction changed over time with regard to its noun types? Our diachronic data from the COHA would allow us to explore that question through the examination of line graphs that show increases or decreases in frequency for different nouns, or different categories of nouns. In this

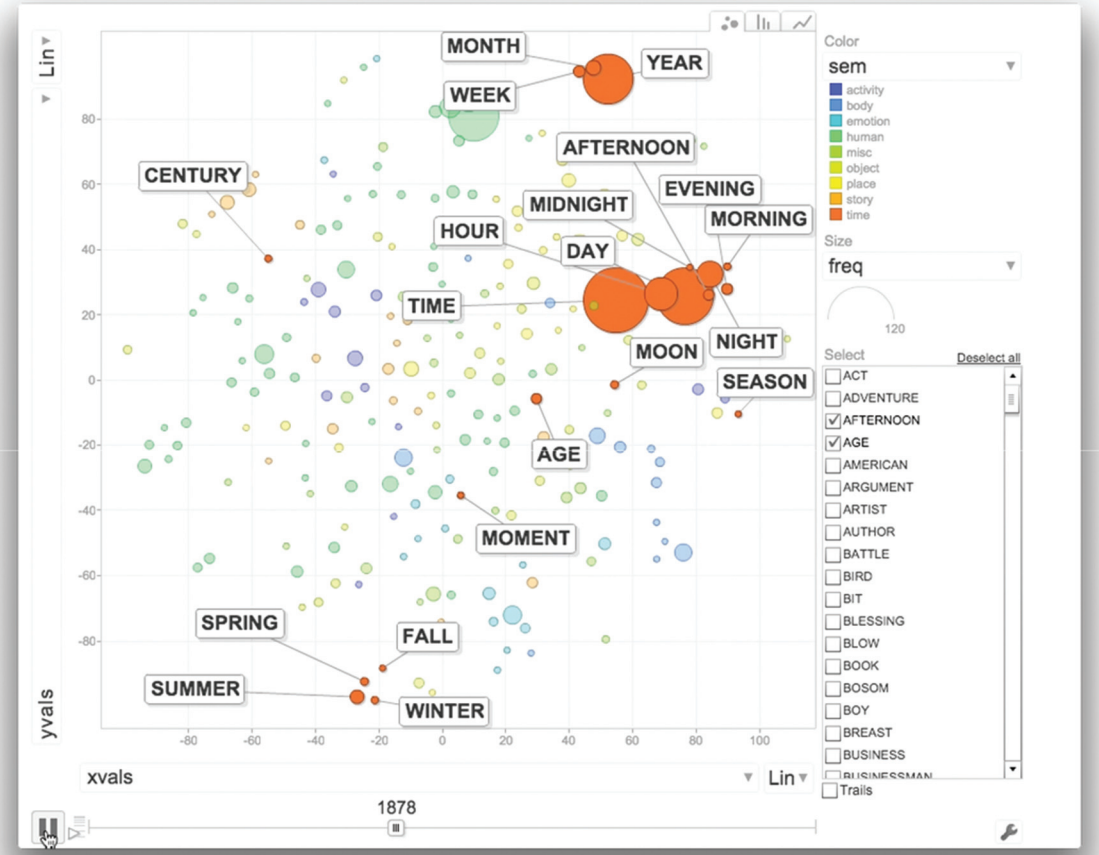
section, we pursue a different way of visualizing that information. By 'animating' the scatterplot of Figure 2 and letting it unfold over time, we represent not only developments in frequency, but furthermore we make it possible to inspect dynamic changes in the semantic spectrum of the construction. Crucially, the animation of the graph does not bring about any movement on the part of the elements on the scatterplot. The relative position of each element is determined by the semantic vector space and thus remains constant over time. What does change is the relative size of each element. Those sizes are determined by our diachronic frequency data. If an element is not attested in a given COHA decade, it disappears from the graph. If it gains in frequency over time, its size increases. Video 1 is a dynamic representation of our data. Its initial state represents the nouns that are attested in the many a NOUN construction in the 1810s. Each noun is represented by a bubble, bubble sizes indicate normalized text frequencies. As in Figure 2, the different colors of the bubbles indicate the different semantic categories.



Video 1: Semantic change in the *many a NOUN* construction.

The single most important observation that can be made from watching the dynamic representation of our data in Video 1 is that throughout the twentieth century, more and more noun types become less frequent and ultimately disappear from our 'petri dish'. The first few decades, between the 1810s and the 1860s, appear to reflect an increase in type frequency, but we would submit that this impression is likely to be an artefact of changes in coverage and composition of the COHA. What is robust is the decline in type and

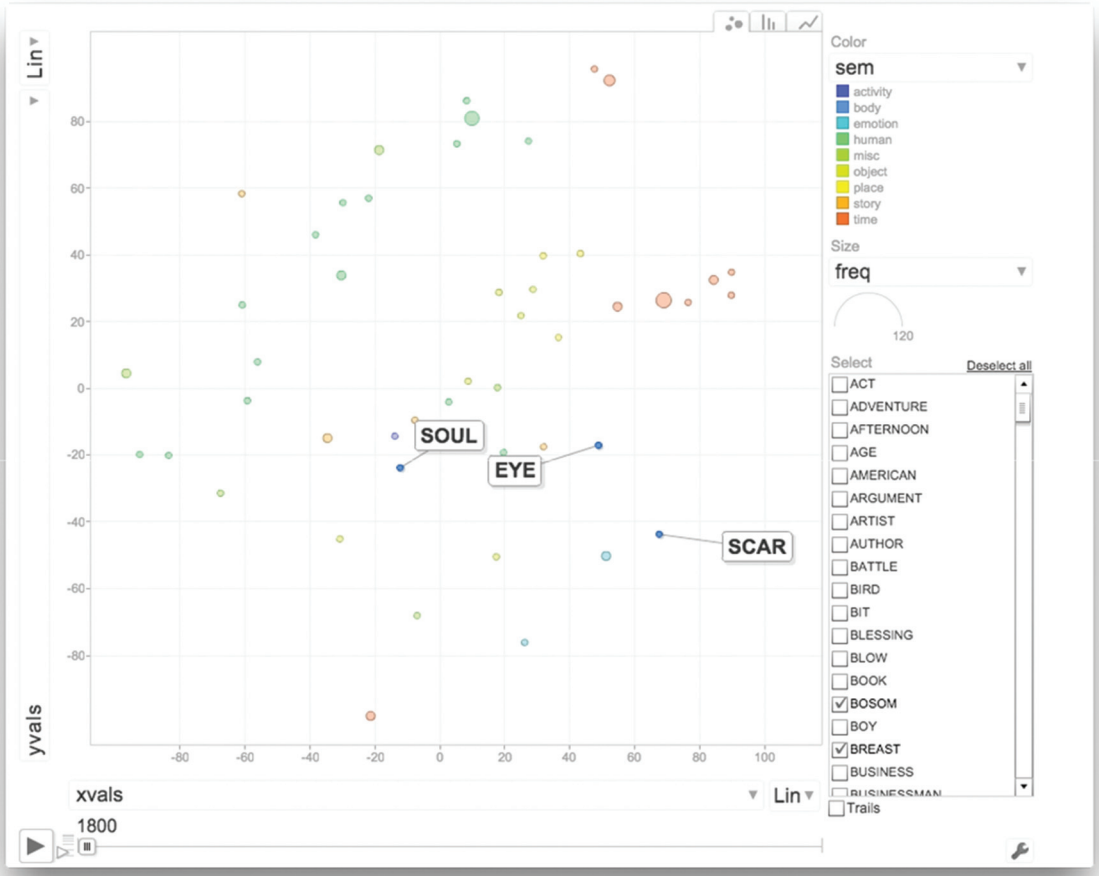
token frequency during the later decades. At the same time, it is evident that the two-dimensional semantic vector space that is shown in Video 1 does not shrink as such. Noun types at the edges of the distribution do persist, even as more and more types disappear. To understand this development in some more detail, it is instructive to look at the development with a focus on individual semantic categories. Video 2 shows the development of nouns that we categorized as time expressions.



Video 2: Change in time nouns in *many a NOUN*.

Time nouns represent the semantic category with the most high-frequency types in the *many a NOUN* construction (*time*, *day*, *year*, *hour*, *night*). At the same time, the type frequency of the category is relatively low, as compared to categories such as human beings. All time expression types that are shown in Video 2 are already attested during the 1830s, their token frequencies increase right up to the 1890s. During the twentieth century, token frequencies recede and more peripheral members of the category (the four seasons, *century*) disappear. Yet, a substantial number of types persists and time nouns continue to stand out as the category with the most frequent types.

A different development can be observed with body part nouns, as shown in Video 3. While also body part nouns peak in their type frequency in the 1830s, their subsequent development differs from time expressions in that they gradually disappear. By the 1970s, only the noun *soul*, arguably a fairly unprototypical body part, remains in the corpus data.

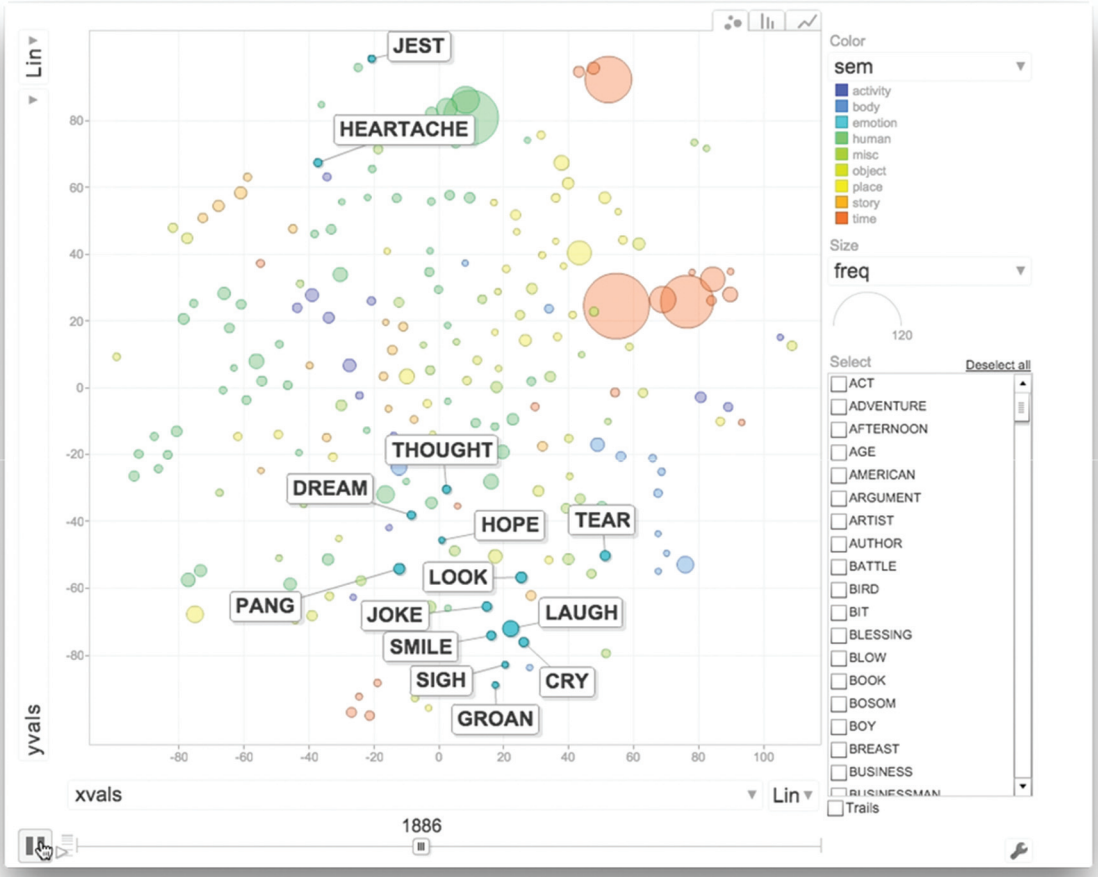


Video 3: Change in body part nouns in *many a NOUN*.

Like body parts, the semantic category of nouns that are related to human emotions shows a near-complete recession. By the 2000s, only the noun *sigh* is found in the COHA data. The disappearance of emotion-related types aligns well with Hilpert's (2012b) finding that other semantic categories come to the fore during the twentieth century (Video 4).

Finally, Video 5 visualizes the development of nouns that denote human beings. In the video, not all types that belong to this category have been marked up, for reasons of better legibility. What is apparent is that the late nineteenth and early twentieth century is characterized by a rich variety of types that recedes in the second half of the twentieth century. Yet, human beings remain the strongest category in terms of type frequency, which again is in line with the result that the *many a NOUN* construction is used predominantly with time nouns and human nouns in twentieth century corpus data.

The overall conclusion that we draw from this case study is that there is a semantic explanation for the persistent productivity of the *many a NOUN* construction. There is empirical evidence that constructions that fall into disuse exhibit a gradual decrease of type frequency (cf. Boyd and Goldberg 2011 on *a*-adjectives; Barðdal 2011 on Icelandic dative subjects, or Rosemeyer 2014 on the replacement of Spanish *ser* plus participle by *haber* plus participle). The *many a NOUN* construction confounds our expectations about this typical trajectory: despite a substantial decrease in its extent of use, it has not receded into a narrow semantic niche. Speakers are still free to use the construction with any noun type they wish to select. Yet, the data reveal that not all noun types are used to equal extents. The figures above have shown

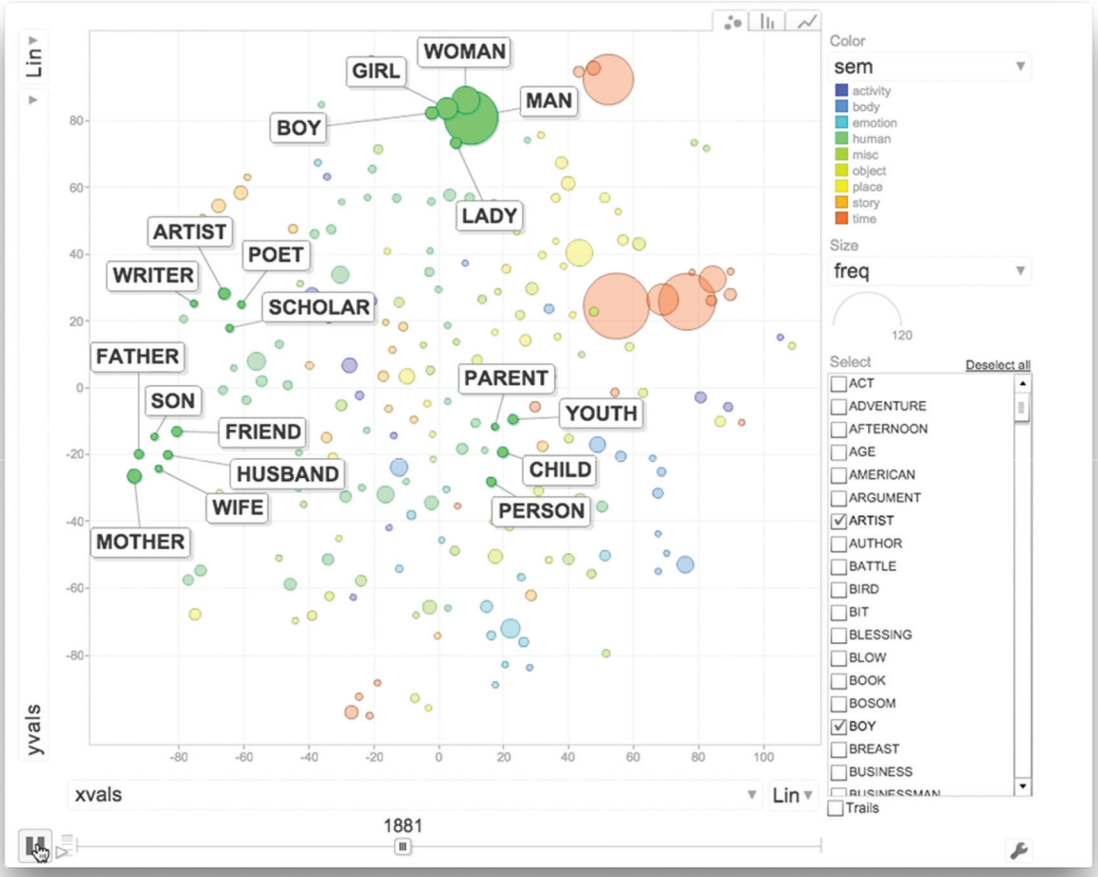


Video 4: Change in emotion nouns in *many a NOUN*.

that while some semantic categories disappear or are substantially weakened, time nouns remain strong, and human being nouns remain strong. What characterizes these nouns is a high degree of semantic generality. Words like *time* or *man* are highly diffuse in their collocational behavior. Furthermore, the category of human being nouns continues to be represented by a sizable residue of semantically diverse types. The result of this constellation is that speakers experience the *many a NOUN* construction as semantically unrestricted. Within the nouns that are regularly encountered in the construction, there is too little semantic overlap for speakers to make out a restriction to well-defined semantic categories.

5 Conclusions

In this paper, we have presented a relatively new visualization tool, motion charts, and we have discussed how it can be usefully combined with semantic vector space modeling in order to study shifts in meaning in grammatical constructions. We have chosen the metaphor of a 'petri dish' to come to terms with dynamic developments in the semantic spectrum of a construction. In our case study, we have observed a recessive development of the elements that populate our semantic petri dish. Yet, this recession did not result in 'empty' areas. We merely observed that the entire area became less densely populated, with some areas retaining a relatively high population density. On the basis of this observation, we have proposed a



Video 5: Change in human being nouns in *many a NOUN*.

semantic explanation for the puzzling fact that the *many a NOUN* construction, despite decreasing type and token frequency, retains its high productivity in Present-Day English.

It goes without saying that visualizations of the kind that we present in this paper are not, in themselves, analyses of linguistic change. Rather, these visualizations provide evidence for or against analyses of linguistic change that the researcher needs to work out qualitatively, on the basis of existing hypotheses or theoretical frameworks. Used in this way, the visualizations can constrain possible analyses and illuminate aspects of change that might have gone unnoticed otherwise. Ultimately, the usefulness of the method depends on the empirical results that future analyses will reveal. We submit that our technique would be suitable for the study of other diachronic phenomena that still wait to be explored. The corpus data and the analytical techniques are readily available, so that it is now up to us as linguists to figure out how they can be put to use in the most fruitful way.

Funding: Schweizerischer Nationalfonds (SNF) (Grant/Award Number: 100015_149176/1).

References

Barðdal, Jóhanna. 2011. The rise of dative substitution in the history of Icelandic: A diachronic construction grammar approach. *Lingua* 121(1). 60–79.

Boyd, Jeremy K. & Adele E. Goldberg. 2011. Learning what not to say: The role of statistical preemption and categorization in "a"-adjective production. *Language* 81(1). 1–29.

Bullinaria, John & Joseph Levy. 2007. Extracting semantic representations from word cooccurrence statistics: A computational study. *Behavior Research Methods* 39(34). 510–526.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 450+ million words, 1990-present. http://corpus.byu.edu/coca (accessed June 2015).

Davies, Mark. 2010. The Corpus of Historical American English (COHA): 400+ million words, 1810–2009. http://corpus.byu.edu/coha (accessed June 2015).

Goldberg, Adele E. 1995. *Constructions. A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Gries, Stefan Th. & Anatol Stefanowitsch. 2004a. Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1). 97–129.

Heylen, Kris & Tom Ruette. 2013. Degrees of semantic control in measuring aggregated lexical distances. In L. Borin, A. Saxena & T. Rama (eds.), *Approaches to measuring linguistic differences*, 361–382. Berlin: De Gruyter.

Hilpert, Martin. 2011. Dynamic visualizations of language change: Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics* 16(4). 435–461.

Hilpert, Martin. 2012a. Die englischen Modalverben im Daumenkino: Zur dynamischen Visualisierung von Phänomenen des Sprachwandels. *Zeitschrift für Literaturwissenschaft und Linguistik* 169. 67–82.

Hilpert, Martin. 2012b. Diachronic collostructional analysis meets the noun phrase. Studying *many a noun* in COHA. In Terttu Nevalainen & Elizabeth Closs Traugott (eds.), *The Oxford handbook of the history of English*, 233–244. Oxford: Oxford University Press.

Hilpert, Martin. 2013. *Constructional change in English: Developments in allomorphy, word formation, and syntax*. [Studies in English Language] Cambridge: Cambridge University Press.

Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.

Huddleston, Rodney & Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

Perek, Florent. 2014, to appear. Vector spaces for historical linguistics: Using distributional semantics to study syntactic productivity in diachrony. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, Maryland USA, June 23–25, 2014 (pp. 309–314). East Stroudsburg, PA: ACL.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985. *A comprehensive grammar of the English language*. New York: Longman.

Ruette, Tom, Dirk Speelman, and Dirk Geeraerts. 2013. Lexical variation in aggregate perspective. In Augusto Soares da Silva (ed.), *Pluricentricity: linguistic variation and sociocognitive dimensions*, 95–116. Berlin: De Gruyter.

Rosemeyer, Malte. 2014. *Auxiliary selection in Spanish. Gradience, gradualness, and conservation*. Amsterdam & Philadelphia: John Benjamins.

Sagi, Eyal, Stefan Kaufmann & Brady Clark. 2012. Tracing semantic change with latent semantic analysis. In Justyna Robynson and Kathryn Allan (eds.), *Current methods in historical semantics*, 161–183. Berlin & New York: Mouton de Gruyter.

Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collostructions: Investigating the interaction between words and constructions. *International Journal of Corpus Linguistics* 8(2). 209–243.

Turney, Peter D. & Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research* 37. 141–188.

Van der Maaten, Laurens & Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of Machine Learning Research* 9. 2579–2605.

Visser, Frederikus Th. 1963. *An historical syntax of the English Language*. vol 1. Syntactical units with one verb. Leiden: Brill.