Chapter 5
Has *Homo economicus* Evolved into *Homo sapiens* from 1992 to 2014: What Does Corpus Linguistics Say?

Yawen Zou and Shu-Heng Chen

5.1 Introduction

Homo economicus has its early roots in the age of classical economics, when economics was dominated by political economy. Mill (1874) described humans as solely caring about wealth possession and making sensible decisions on their own behalf. These characteristics of *Homo economicus* then evolved over time, especially after Walrasian general equilibrium analysis eventually superseded political economy, when *Homo economicus* became super-rational and infinitely smart, and human decisions were neither dictated nor influenced by emotions. The use of a mathematical optimization framework which was heavily borrowed from operations research, and was developed from the 1940s to the 1960s, laid the analytical foundations of such human behavior and hence helped this paradigm to evolve.

In addition to economics, *Homo economicus* has also been introduced to other sister disciplines in the social sciences, a spread generally known as *economic imperialism*. However, not all social scientists can embrace *Homo economicus* as economists did; in fact, a negative side of economic imperialism is that economics became more alienated from her sister disciplines, and economists sometimes considered economics to be a "brother" of the physical sciences, including mathematics. The communication between economics and other social sciences turns out to be much weaker than an outsider can possibly imagine. Humans, in the eye of

Y. Zou (✉)
Center for Biology and Society, Arizona State University, Tempe, AZ, USA

The Chinese University of Hong Kong, Shenzhen, China
e-mail: yzou20@asu.edu

S.-H. Chen
AI-ECON Research Center, Department of Economics, National Chengchi University, Taipei, Taiwan

S.-H. Chen (ed.), *Big Data in Computational Social Science and Humanities*, Computational Social Sciences, https://doi.org/10.1007/978-3-319-95465-3_5

most economists, behave in a way that is rather peculiar in so far as other social scientists can perceive. This difference, well characterized by the increasing gap between what people *ought to do* and what they *actually do*, created a degree of uneasiness between economists and other social scientists. This gap also gave rise to the distinction between "good economics" and "bad economics"; a world filled with *Homo economicus* cannot possibly experience a financial crisis (McDonald 2009).

To narrow the gap and lessen this uneasiness, economists began to work with an alternative *Homo*, namely *Homo sapiens*. Over the past few decades, behavioral economics has illuminated the role of the social, cognitive, and emotional aspects of humans in economic decision-making (Dohmen 2014; McDonald 2009), and hesitates to characterize humans as self-interested, utility- or profit-maximizing machines. In his popular book *Not Just for the Money*, Frey (1997) distinguished intrinsic motivation from extrinsic motivation (the pecuniary motivation) and highlighted the neglected importance of the former in understanding human choices and behavior. Through his life-time devotion to the study of cooperation and altruistic behavior, Nowak and Highfield (2011) argued that both altruism and cooperation were both significant in terms of the survivability of an individual or even the whole of mankind.

Unlike *Homo economicus*, *Homo sapiens* is what Herbert Simon coined as being *boundedly rational*, and is constrained by limited memory and computational capacity. The decisions made by such humans are, therefore, further constrained by a great variety of social settings and emotional drives. Based on these developments, Thaler (2000), a leading behavioral economist, predicted that, in the world of economics, *Homo economicus* would eventually "evolve" into *Homo sapiens*. Since more than 15 years have now passed since Thaler made his insightful prediction, it seems to be a good time to evaluate whether Thaler's prediction is correct in the sense that the research paradigm characterized by *Homo economicus* has been gradually replaced by the paradigm characterized by *Homo sapiens.*

The underlying assumption of this study is that the changes in these two research paradigms are reflected by a change in the language used by economists. Research articles are an important genre of the economists' discourse, and this era of big data has bombarded us with a huge number of readily available research articles. When "manpower" is too limited to surf over such an extensive "ocean" of studies, adopting an automated or computational approach toward a corpus built from a myriad of research articles has become increasingly practical (Biber et al. 1998). A corpus is a collection of textual data, and it appears in the form of either written languages or spoken languages. Written language can comprise texts retrieved from online websites, like Twitter, Amazon reviews, eBay product descriptions, etc., or it can be retrieved from research articles, as in our case (Knight et al. 2015).

Many humanists have used the techniques of corpus linguistics to gain knowledgeable insights into a particular domain that interests them. Corpus linguistics was initially used by linguists to study language patterns, but it was then adopted by scholars for different areas of investigation. Scholars can apply the corpus linguistic approach to study any section of the research articles, such as the abstract, discussion sections, and methodology sections (Flowerdew 2015). The approach has also been

used to study the changes in concepts and ideas. For example, Pumfrey et al. (2012) analyzed the changes in the meaning of the word "experiment" from early English books, using both the corpus linguistics approach and the manual approach, i.e., a traditional close reading of randomly selected texts. It was found that the former approach proved to be more efficient than the latter.

This chapter is, to the best of our knowledge, a pioneering application of corpus linguistics to economics. We shall first address the issue of how to build a corpus that can represent the economics literature and introduce a tool for the corpus linguistic analysis. Second, to have a concrete illustration, we shall use the stemmed word, "cognit*," to demonstrate the analysis conducted in this paper. The same kind of analysis will be applied to other 100 keywords, which are picked by the machine and a human expert. Third, we further divide the whole corpus into two periods, namely before and after the year 2000, and carry out a co-word network analysis to examine any discernible changing pattern. Finally, we summarize the main findings of our study by remarking on its current limitations and the opportunities for future research.

5.2 Methods

To trace the evolution of a field, one approach that economists are very familiar with is bibliometrics, a methodology that has long been used to trace the specific dynamics of a discipline (De Bellis 2009). The approach demonstrated herein, however, is atypical; it is motivated by the recent applications of corpus linguistics and network analysis in studying the history of ideas. Since this approach is not familiar to most economists, providing a brief background should be useful.

To begin with, we wish to point out that neither *Homo economicus* nor *Homo sapiens* refers to a well-defined research discipline or methodology. This lack of articulation makes the conventional bibliometric approach hardly applicable to tracing the development of the two paradigms. In fact, what really interests us is whether the idea of *Homo sapiens* has spread widely in different territories of economics, and is not just confined to some specific fields, such as behavioral economics, health economics, or labor economics, or to a specific methodology, such as experimental economics, computational economics, or neuroeconomics. In other words, the framework that appears to be more suitable for us is the *history of ideas* (Lovejoy 2011).

There is a well-established area called the history of economic ideas, also known as the history of economic thought, to which little attention has so far been paid (Blaug 2001). As in the study of other histories, the methodology used for the study of the history of economic ideas is mostly narratives-oriented. It is usually based on the study of a number of magnum opuses and influential articles. Normally, we can learn from these painstaking studies and infer when an idea of interest was first introduced, why it was needed, how it was formulated, and how it then evolved, with each incidence or change supported by major references. Needless to say, this classical narrative approach will continue, since, as one may rightly argue, only man

can read and think. However, these studies are time-consuming, and researchers using this approach will find it harder to keep up with the rapid growth of the literature. Furthermore, no one can be assured of what may be missing given the voluminous amount of literature.

The recent information and communication technology revolution provides us with a new division of labor between humans and machines that makes various novel forms of cooperation between humans and machines possible. One concrete example is digital humanities, the application of digital tools to studying humanities (Schreibman et al. 2008). Not only has the second wave of the Industrial Revolution provided us with new methods, but it has also given us new ontologies and epistemologies. We might then ask: what is the history of ideas, or even more fundamentally, what actually is an idea?

Without losing generality, let us assume that an idea can be presented in the form of written texts, which are in turn collections of words or symbols. Of course, this collection is not just a monotone enumeration of words or a fragmental list of symbols; rather, these words and symbols are embedded within a specific structure, so that they are coherently connected. This description puts the history of ideas in a representation of evolving networks. In this representation, ideas may behave as a biological entity and can be treated as a research object in a complex adaptive system (Simon 1962). A machine may not be able to read and think on the works and ideas found in Shakespeare's plays, but it can help us to dig out the possible networks (structures, patterns) within those works. These extracted patterns or structures may not be complete, but humans can decide how incomplete they are and how they can be made complete. This human–machine interaction can, therefore, fundamentally change our perception of the history of ideas, from its ontology, to its epistemology and methodology.

This chapter, to the best of our knowledge, is the first study in the history of economic ideas that combines the corpus linguistics approach with the co-word network analysis. In the following, we will begin with the basic level of analysis, i.e., a word. A word serves as a node in a network. Hence, being a fundamental unit, it should be treated as the first step toward a fully-fledged study of the history of ideas. We then trace words in a network to see their relationships.

5.2.1 Building a General Economics Corpus

In corpus-based research, the primary task is to build a corpus that is sufficiently representative of a research field that is of interest. As mentioned above, since our interest is to examine how the idea of *Homo sapiens* has been accepted by economists at the global (the mainstream) level, the corpus of this study should be built upon the articles published in mainstream economics journals. Characterizing what mainstream journals are is straightforward. The criterion that we have adopted here is based on ratings. There are many different ratings available. However, since they do tend to overlap quite substantially, we can simply follow one of them. In this study, we follow the rating provided by Kalaitzidakis et al. (2011).

Kalaitzidakis et al. (2011) analyzed the number of citations for the previous 10 consecutive years for articles published between 2003 and 2008 in a number of journals. We pick this rating because their survey covers a longer horizon, as opposed to other rankings, such as the rating in Thomson Reuter's Journal Citation Report, which is based on the Web of Science database[1] and covers only 2 years of citations, and the SCImago journal ranking, which is based on the Scopus database[2] and also covers only 2 years of citations. Since the mainstream should be reasonably enduring, we therefore choose a ranking that is less sensitive to time.

We took the list of the 50 top-ranking journals. Most of these journals, such as the *American Economic Review*, are relevant to almost all major fields in economics, but some are either too domain-specific or too technical. The latter kind of journals may have little relevance to either of the paradigms, be it *Homo economicus* or *Homo sapiens*. As including the journals of the latter kind could blur our research focus, we decided to remove those journals. Since this decision inevitably involved a degree of discretion, we only took a mild cut, i.e., 8 journals out of the 50. Table 5.1 presents a list of the 42 journals selected.

Table 5.1 The 42 mainstream economic journals

Index	Journal	Index	Journal
1	Am Econ Rev	22	J Econ Growth
2	Qjecon	23	J Hum Resour
3	Econometrica	24	J Econ Dyn Control
4	J Polit Econ	25	J Econ Behav Organ
5	Rev Econ Stud	26	J Health Econ
6	J Econ Theory	27	J Appl Econom
7	J Public Econ	28	Brookings Pap Eco Ac
8	Econ J	29	World Bank Econ Rev
9	J Econ Perspect	30	Econ Theor
10	J Int Econ	31	Scand J Econ
11	J Econ Lit	32	Oxford Econ Pap
12	J Financ Econ	33	Can J Econ
13	Eur Econ Rev	34	Econ Inq
14	Rand J Econ	35	Econ Policy
15	Int Econ Rev	36	Int J Ind Organ
16	J Eur Econ Assoc	37	Public Choice
17	Game Econ Behav	38	J Law Econ
18	Econ Lett	39	World Dev
19	J Dev Econ	40	J Law Econ Organ
20	Rev Econ Dynam	41	J Ind Econ
21	J Labor Econ	42	Labour Econ

[1] http://wokinfo.com/products_tools/analytical/jcr/

[2] http://www.scimagojr.com/journalrank.php?area=2000

Table 5.2 Number of abstracts and word counts of the sub-corpus in each year

Year	No. of articles	No. of abstracts	Word counts
1992	2546	1260	118,795
1993	2459	1293	125,616
1994	2444	1604	152,960
1995	2453	1686	163,976
1996	2715	1805	176,908
1997	2603	1794	177,326
1998	2676	1876	193,881
1999	2485	1838	193,951
2000	2575	1886	201,736
2001	2714	2017	218,614
2002	2708	2087	223,984
2003	2836	2238	241,386
2004	2821	2262	245,601
2005	2734	2283	251,377
2006	2760	2386	261,590
2007	3039	2645	297,015
2008	3291	2899	322,744
2009	3042	2657	308,338
2010	3046	2701	308,201
2011	3115	2718	317,283
2012	3528	3150	350,411
2013	3490	3135	374,583
2014	3397	3065	371,259
Total	65, 477	51, 285	5,597,535

We then built a corpus of over five million words using the self-developed Python code to retrieve the abstracts of research articles published from 1992 to 2014 for the selected 42 mainstream economics journals. These included the *American Economic Review, European Economic Review, Quarterly Journal of Economics*, etc. We downloaded the metadata of those articles from the *Web of Science* database. Table 5.2 provides the size of our corpus year by year, from 1992 to 2014. Over the years, the annual corpus size has increased from 118,795 words in 1992 to 371,259 words in 2014.

Thaler made his prediction about *Homo sapiens* in 2000. That year is roughly in the middle between 1992 and 2014, so the time frame employed in this study captured 8 years prior to his prediction and 14 years after his prediction. We picked the year 1992 as a starting point instead of earlier years for a precise reason: the abstracts of most journals were not available in the Web of Science database until 1992. For example, in 1990, the 42 journals of choice published a total of 2169 articles, but only five had abstracts deposited in the database. In 1992, these 42 journals produced 2546 articles, and 1269 of them had abstracts. In total, from 1992 to 2014, the 42 journals produced 65,477 articles, and 51,285 of them had abstracts.

For the numbers of published articles, their available abstracts, and the word counts of the sub-corpus for each year, the reader is referred to Table 5.2.

To facilitate a diachronic study of the language patterns in the abstracts of these research articles, we grouped the abstracts published in the same year into a single file to have a total of 23 files, with each file corresponding to one of the years between 1992 and 2014. We then analyzed the frequency of words in each year.

5.2.2 *Analyzing a Selected Word List Using WordSmith*

After building the corpus, we identified a set of keywords to study. A paradigm shift can involve the deletion of some old ideas (represented by words) and the addition of some new ideas. The purpose of having a set of keywords is to identify the deleted and the added ideas. The practice of obtaining keywords needs to be either manually provided by experts or to be automatically generated by computer or by both. We followed the third route by first having the computer search for a list of frequently appearing words. WordSmith is one of the mostly widely used tools in corpus linguistics (Scott 2001).

We first used the WordSmith *Keywords* tool to identify a list of keywords. These words automatically satisfy the necessary condition of being the key, but they are not sufficient. As one could well expect from Zipf's law, many very frequently appearing words are just articles, pronouns, and modifiers (Zipf 1949). Hence, in the second stage, we removed these "nuisances," and used our domain knowledge to select a list of 101 words as the keywords. These are presented in Table 5.3.

Table 5.3 The 101 words that are related to the two paradigms

Category	Words or stemmed words
Homo sapiens (48)	Adaptive, affect*, agent-based, ambiguity, anthropolog*, attitude*, behavior*, bias*, chaotic, cognit*, complexity, cooperat*, cultur*, darwin*, decentraliz*, donor*, emergence, emotion*, ethnic*, evolut*, experiment*, happiness*, heterogen*, imperfect, incentive*, instabilit*, interact, intuition, laboratory, neighbor*, network*, norm, optimism, overconfidence, psycholog*, religi*, satisfaction*, selection*, social*, socio*, stimulus*, subjective, trust, unbiased, uncertaint*, uninformed, well-being*, wisdom
Homo economicus (53)	Algorithm*, anticipated, approximation*, axiom*, centralization*, complementarity, computational, consensus*, convergence*, counterfactual, difficult*, efficien*, equilibr*, expect*, experience*, feedback, first-best, first-order, forecast*, free-riding, habit, idiosyncratic*, informational, intertemporal, leader*, logic, maxim*, minim*, model*, motivated, noncooperative, normative, optimal*, optimiz*, optimum, perceived, perfect, plausible, predict*, preference*, profitability, random*, rational*, reason*, satisfi*, search, selfish, simulat*, stationary, steady-state, tractable, tradeoff, utilit*

It is hard to give a specific account for how each of these words was selected. By and large, we roughly taxonomized the words into three groups: the first group of words facilitated the expression of the idea, *Homo economicus*; the second group of words facilitated the expression of the idea, *Homo sapiens*; and the third group of words was neutral to both paradigms. Two remarks need to be made here, however. First, needless to say, this taxonomy makes a bold assumption for language, since neither meaning nor function is context free. The justification that we make is at best a first-order approximation, and bringing the context and embeddedness together is necessary if our initial analysis can indicate that this is an interesting direction for future research. Second, even though we can assume such a taxonomy, any manual classification of the words into the above three categories might still suffer from a certain degree of arbitrariness. In other words, different "experts" can come up with different taxonomies. We certainly cannot exclude such possibilities, but we can assume that such differences are mostly of a secondary or minor degree and hence will have little effect on the results.

With the above two qualifications in mind, we identified the first and the second group of words. Of the 101 words we picked, 48 words were related to the *Homo sapiens* paradigm (the first group) and 53 words were related to the *Homo economicus* paradigm (the second group). A quick look at Table 5.3 shows that many of the words in the former relate to psychology, sociology, or other social sciences. For example, "cognit*" and "emotion*" were chosen in light of Thaler (2000): "Economists will study human *cognition*," and "*Homo economicus* will become more *emotional*" (Ibid, p. 137 and p. 139, respectively).

Note that in Table 5.3, for some words, we used the stemmed word and placed a * symbol after the stemmed word. The * symbol enables WordSmith to search for any derivatives of the stemmed word. If we take "cognit*" as an example, WordSmith will count the occurrences of "cognitive," "cognition," "cognitively," "cognitivity," etc. By taking "anthropolog*" as another example, WordSmith searches not only for "anthropology," but also for "anthropological," "anthropologist," etc.

We used the WordSmith *Concord* tool to analyze the frequencies of these 101 words year by year from 1992 to 2014. *Concord* indicates all references for each single word in our corpus and can show the *concordance lines*, which enables us to look closely at how each word is used in a sentence. It can also show the frequencies of a word in any given text (Scott 2001). Each of our documents is indexed by year, and the frequencies of a word across all documents can be arranged to appear in the form of a time series.

After a set of keywords was determined, the application of the corpus linguistics to the study of the paradigm shift in economics could be formulated into a *trending hypothesis*. The trending hypothesis indicated that the frequencies of words that are relevant to the paradigm of *Homo economicus* have a tendency to decline over time, whereas the frequencies of words relevant to the paradigm of *Homo sapiens* tend to increase over time.

5.2.3 A Linear Regression Model to Analyze Word Frequencies

A standard way to see whether a word has a trend is to run a simple linear regression and to regress the word frequency (the regressor) against time (the regressand) (Bianchi et al. 1999). The trend can then be determined by the slope of the regression line. A positive slope indicates an upward trend, while a negative slope indicates a downward trend. At the same time, the R-squared value is also calculated for each word, which means that the proportion of variation explained by the model, or simply, how good the model fits the real data, can also be known. We used Python codes to construct the regression model.

In this study, we also checked the *p*-values for the trend. A rule of thumb applied here is that the trend is statistically significant if the *p*-value is lower than 0.05. Words without a significant trend were removed from the modified word list, and our subsequent analysis only focused on words with a statistically significant trend. For those words which had a *p*-value lower than 0.05, we plotted the estimated trend (their slope) on the *x*-axis and the corresponding R-squared value on the *y*-axis (see Figs. 5.4 and 5.5 below).

5.2.4 Co-word Network Analysis of the Economics Literature Before 2000 and After 2000

We are not only interested in the trend of the individual words, but also in how the relationships among words changed. We carried out a co-word network analysis based on the abstracts published before 2000 and the abstracts published after 2000, respectively, and compared how the two networks differed. The co-word network analysis can be used to discover the relationship between words by exploring which words co-occur with other words (He 1999). We used the tool referred to as *ConText* to implement the co-word network analysis (Diesner 2014), and examined the relationships among the 101 words mentioned above.[3]

5.3 Results

In this section, we first use the stemmed word "cognit*" as an illustration to show how each stemmed word is analyzed in this chapter (Sect. 5.3.1). Then, in Sect. 5.3.2, we show the overall frequency trend for both the *Homo sapiens* words and *Homo economicus* words. Finally, in Sect. 5.3.3, we offer a microscopic

[3]Since only WordSmith can recognize stemmed words, a.k.a., the "*" symbol, and *ConText* cannot, we removed the * symbol and changed the stemmed words to normal words when using the tool *ConText*.

interpretation that explains why some words exhibit an upward trend and some words a downward trend. We propose four reasons that could explain the upward trend observed for some *Homo sapiens* words.

5.3.1 Concordance for the Stemmed Word "Cognit*"

A typical result is demonstrated below using the stemmed word "cognit*." Figure 5.1 shows the first 25 of 456 sentences that include the derivatives of the word "cognit*." This figure provides us with a close examination of the context in which this keyword of interest is situated. For example, the second entry involves Frederick's (2005) discussion of the relationship between cognitive reflection and decision-making.

WordSmith can show how many hits a word has in each year and the normalized frequency of that word in every 1000 words. Figure 5.2 visually shows that the derivatives of "cognit*" have only a few occurrences (see the column "Plot") before the year 2000, but more occurrences in later years.

Figure 5.3 is a visual representation of the normalized frequency of "cognit*." The trendy property of this stemmed word is represented by a fitted regression line as shown in Fig. 5.3. Despite a degree of fluctuation, there is clearly an upward trend in terms of its use. The accompanying regression line shows that this linear model

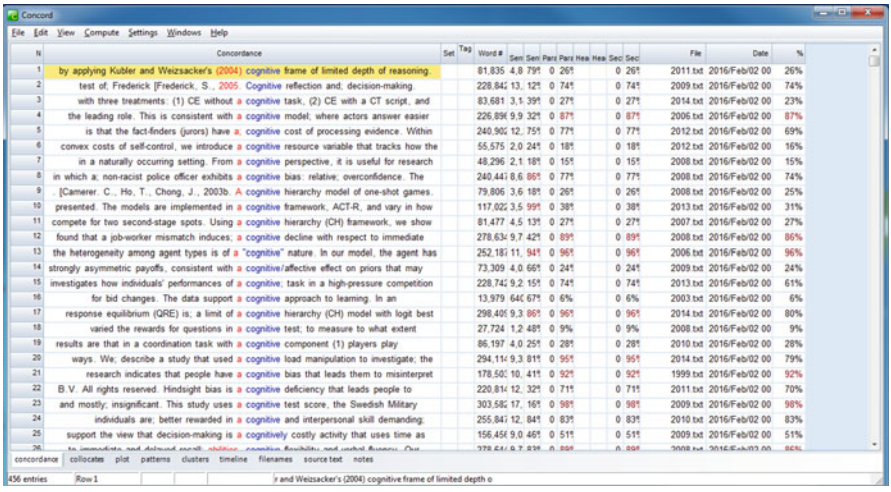

Fig. 5.1 The concordance lines for the stemmed word "cognit*." This page was generated by the WordSmith *Concord* tool. The first column is the index of the concordance line. The second column is the concordance line, which shows the sentence that includes the derivatives of the stemmed word "cognit*." In the bottom left corner of the figure, it is stated that there were 456 entries, meaning that in our corpus, the derivatives of the stemmed word "cognit*" appeared 456 times

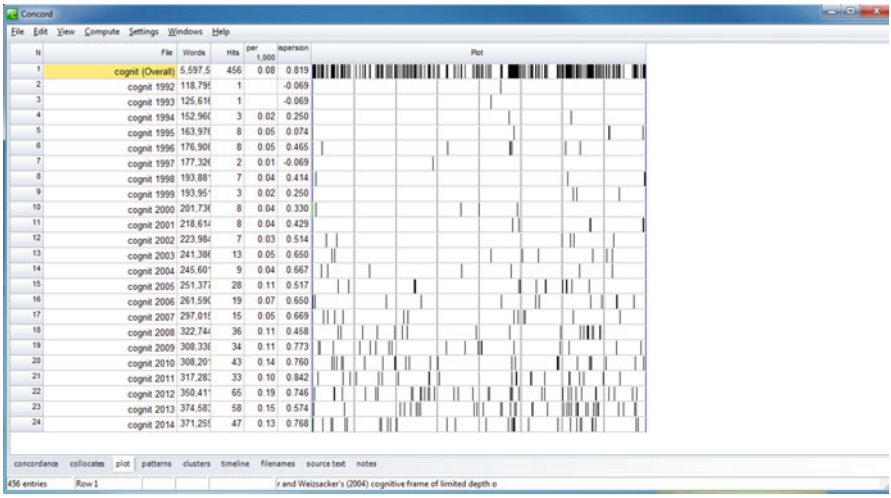

Fig. 5.2 The concordance plot for the derivatives of "cognit*." This figure was also generated by the WordSmith *Concord* tool. The "File" column denotes the files corresponding to each year. For example, "Cognit 1992" refers to the file which includes all abstracts published in 1992. The "Words" column shows how many words there are in the respective file. The "Hits" column shows how many times the derivatives of "cognit*" appear in a file. The "Per 1000" column shows the normalized frequency of the derivatives by averaging their frequencies of appearance for every 1000 words. The "Plot" column shows the position where the derivatives appear in a file by reconfiguring the text as a strip

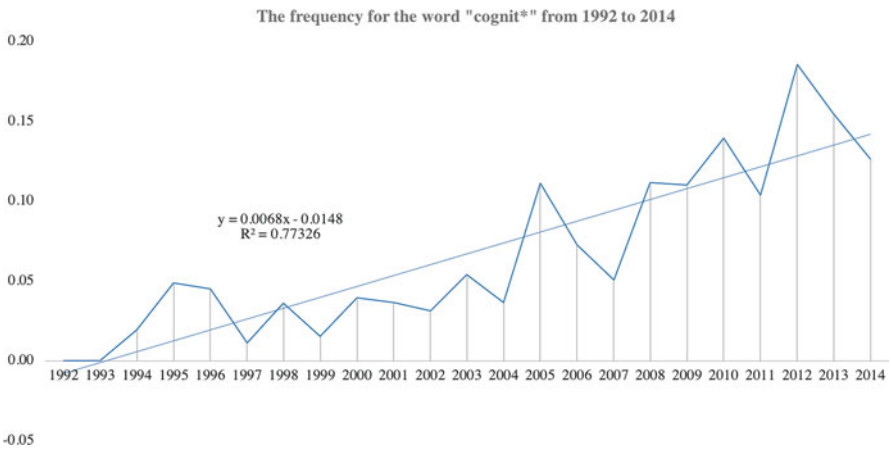

Fig. 5.3 The normalized frequency of the derivatives of "cognit*" increases with time (the curved line) and the fitted regression (the straight line). The *x*-axis denotes the year, and the *y*-axis denotes the normalized frequency

is a good fit because the fitted line has an *R*-squared value of 0.773. The slope of the linear regression line, 0.0068, is positive, and has a *p*-value of 3.30453E-08, much smaller than 0.05, which indicates that the trend is both statistically significant and positive.

5.3.2 A Macroscopic Examination of the Trendy Keywords

The linear regression analysis as illustrated for the stemmed word "cognit*" was applied to all the other 100 keywords. For brevity, in the rest of this chapter we use "word" to indicate not only a word but also a stemmed word. Out of a total of 101 words, 12 were not found to be trendy (the p-value of the estimated slope being higher than 0.05). They are therefore not considered in our further analysis. The remaining 86 words are demonstrated in an *x*–*y* plot in Figs. 5.4 and 5.5, where the *x*-axis denotes the trending coefficient and the *y*-axis denotes the associated R-squared value.

To make sense of these results, we first considered an extreme, but ideal, pattern of the trending hypothesis. Specifically, we expected to see that those words associated with the *Homo sapiens* paradigm had a positive slope, while those words associated with the *Homo economicus* paradigm had a negative slope. However, this is a rather ideal situation, as the methodological restrictions or the violations of the simplified assumptions that occurred previously could all cause some degree of deviation. Nevertheless, we were still able to ask whether the majority of words did in fact behave in line with our expectations.

Table 5.4 shows that, for the *Homo sapiens* words, this was roughly the case: 81.4% (the majority) of the words (35 out of 43) had an upward trend, and only 18.6% (the minority) of the words (8 out of 43) had a downward trend. For the *Homo economicus* words, we see a similar, but a less pronounced, pattern: 65.2% (a mild majority) of the words (30 out of 46) had a downward trend, while 34.8% (a mild minority) of the words (16 out of 46) had an upward trend.

Hence, from a macroscopic viewpoint, the paradigm of *Homo sapiens* did gain momentum over time, and that gain, to some extent, can be translated into the gradual decline of the paradigm of *Homo economicus*. From this observation, we can conclude that Thaler's prediction is largely correct. In fact, considering that our period of observation is not sufficiently long, one may expect this predicted shift to be still ongoing.

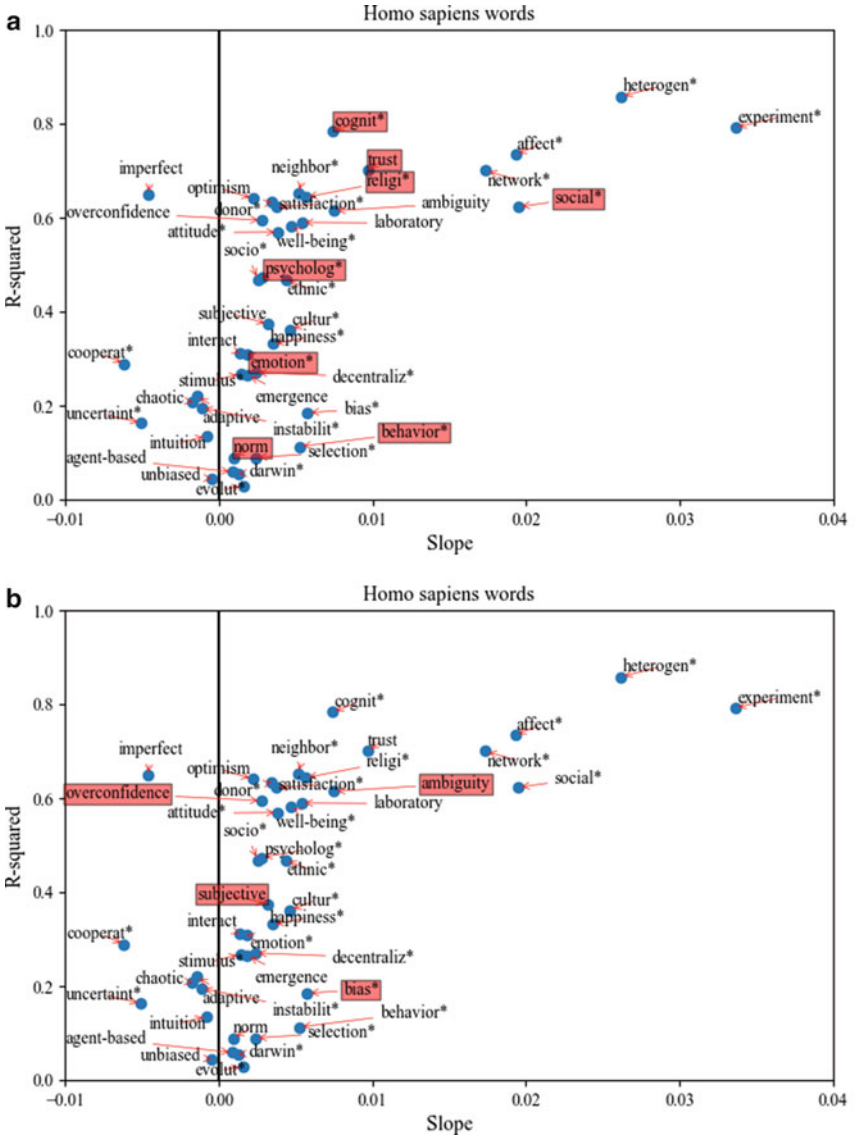

Fig. 5.4 (**a**) The trending coefficient (*x*-axis) and the R-squared value (*y*-axis) of the first group of words that relate to the paradigm of *Homo sapiens*. (**b**) The trending coefficient (*x*-axis) and the R-squared value (*y*-axis) of the second group of words that relate to the paradigm of *Homo sapiens*. (**c**) The trending coefficient (*x*-axis) and the R-squared value (*y*-axis) of the third group of words that relate to the paradigm of *Homo sapiens*. (**d**) The trending coefficient (*x*-axis) and the R-squared value (*y*-axis) of the fourth group of words that relate to the paradigm of *Homo sapiens*

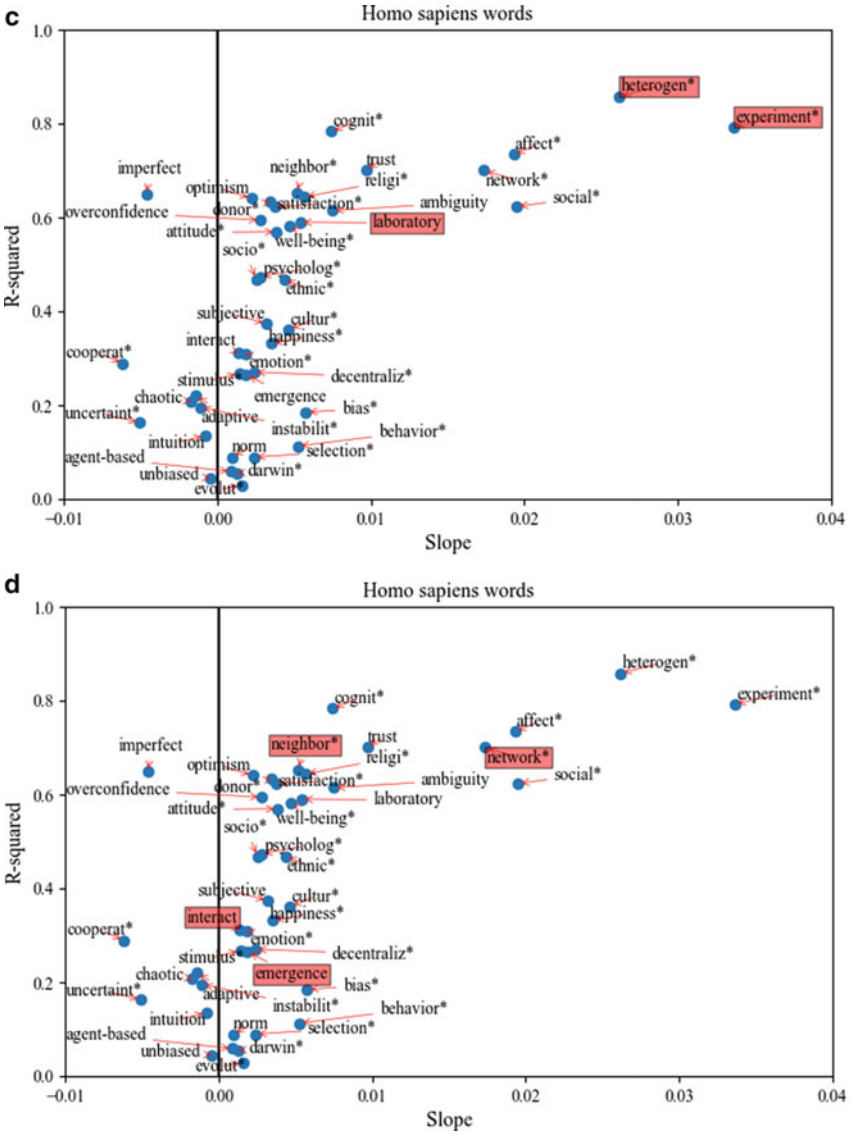

Fig. 5.4 (continued)

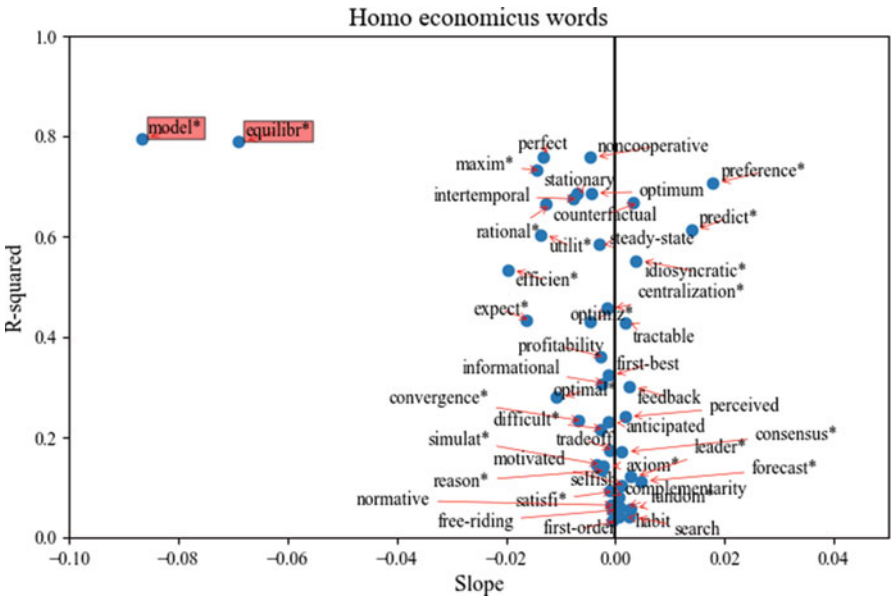

Fig. 5.5 The trending coefficient and R-squared value of words related to the *Homo economicus* paradigm

Table 5.4 The words with a positive slope and a negative slope

	Homo sapiens	*Homo economicus*
Total number of words	48	53
Words with a p value <0.05	43	46
Words with a positive slope	35	16
Words with a negative slope	8	30

5.3.3 *A Microscopic Examination of the Trending of Words in the Entire Corpus*

In addition to the macroscopic perspective, it is interesting to provide a microscopic examination of words. We first, in Sect. 5.3.3.1, discuss the words related to the *Homo sapiens* paradigm, and then, in Sect. 5.3.3.2, the words related to the *Homo economicus* paradigm. We are not only interested in words that fit our hypothesis, but are also interested in the words that deviate from our expectations.

5.3.3.1 *Homo sapiens* Words

In this section, we first examine the 35 *Homo sapiens* words that have a positive slope. It will, however, be hard and even fragmental to elaborate on them individu-

ally. Some words were conceptually connected, so we grouped them and elaborated on a group of words. We identified four groups of words that had a positive slope. We then addressed how each of these groups can help *Homo sapiens* get established.

The first group of words, highlighted in Fig. 5.4a, is related to other sister disciplines in the social sciences, for example, "psychology*," "cognit*," "emotion*," "social*," "trust," "norm," "religi*," "behavior*," etc. All these words, when put together, indicate that the idea of *Homo sapiens* is the consequence of long-standing interdisciplinary integration, by which economists have accepted ideas from psychology, sociology, anthropology, ethnics, and cultural studies.

Thaler (2000) predicted that "*Homo sapiens* will begin losing IQ" and "will be a slow learner" (Ibid, p. 134, 135, respectively). The second group of words, highlighted in Fig. 5.4b, including words such as "bias*," "ambiguity," "overconfidence," and "subjective", are words that represent the (cognitive) constraints of humans, in terms of IQ, which may in turn influence their decision-making. "Overconfidence," for example, has a steep slope, which basically reflects increasing general concerns that economists have for their economic man: economic miscalculation is not an exception, but a rule.

The third group of words, highlighted in Fig. 5.4c, is related to the heterogeneity of humans, as exemplified in the following words: "heterogen*," "laboratory," and "experiment*." The paradigm of *Homo sapiens* assumes that agents are non-trivially different or that they are heterogeneous. To harness their behavior, one cannot just count on the analytical models, but should use the "laboratory" approach by running "experiments." The third group of words supports this claim.

The fourth group of words, highlighted in Fig. 5.4d, are "network*," "neighbor*," "interact," and "emergence." These words are all concepts related to the complexity of economic behaviors. Complexity science studies complex systems, in which many parts interact with each other and conceptually form a network. During the past three decades, complexity economics has emerged as a field that treats economic agents as constantly interacting with each other and changing their behavioral rules or strategies along the course of these interactions (Arthur 2014). The macroscopic patterns of humans or a complex system are not just a linear summing up of individual components, but often have properties not seen at the individual level (the emergent properties).

After going through the four groups of well-expected or justifiable trendy words for the paradigm of *Homo sapiens*, it is also interesting to see those *Homo sapiens* words that actually exhibit an opposite trend. There are seven such stemmed words. Several of them, including "adaptive," "chaotic," "imperfect," and "uncertain*," are words that were already used by heterodox economists well before the 2000s, when such economists attempted to challenge the stringent assumptions of neoclassical economics. For example, they used adaptive behavior to question or challenge rational behavior, used chaotic dynamics to challenge the trivial and uninteresting linear dynamics, and used the environment characterized by true uncertainty and the resultant imperfect information to address the difficulty of rational calculation and the implausibility of an expected utility maximization framework. What, then,

caused the importance of these words to decline over time? We believe that this is an open question that can only be answered by further research.

5.3.3.2 *Homo economicus* Words

In Sect. 5.3.2, we have seen the increasing frequencies for the majority of the *Homo sapiens* words, and the decreasing frequencies for the (mild) majority of the *Homo economicus* words. After analyzing the positive trend of the former, in this subsection, we examine the negative trend of the latter. There are 30 *Homo economicus*-related words with a negative slope. In Fig. 5.5, we have highlighted the two that noticeably stand out, namely "model*" and "equilibr*." These two, as compared to many others, have both a steep negative slope (a sharp declining rate) and a larger *R*-squared value.

The sharp decline in the normalized frequency of the word "equilibrium" is interesting but not surprising. The concept of equilibrium is a centerpiece of the paradigm of *Homo economicus*, and it goes hand in hand with "rationality." *Homo economicus* is assumed to be rational as if it is mathematically "optimizing" a well-specified objective function. Not only does this well-articulated behavioral formulation help characterize what an equilibrium is, but it also helps provide a solution to the model, which is normally characterized as the "steady state(s)" of the model. Once the steady state is determined, the remaining issue left for the dynamic analysis will be the path "converging" to the steady state.

Homo sapiens, on the other hand, is assumed to be boundedly rational. The mathematical description of the behavior of *Homo sapiens* is far from just optimizing and, worse, is normally not homogeneous and unique. Putting all of these together makes the model-solving a daunting task if not impossible. Under these circumstances, the equilibrium is no longer operational for the model. Neither is the "steady state," nor is "convergence". In Fig. 5.5, we see that in addition to "equilibrium," "rational," "optimal," "steady-state," and "convergence" all exhibit a declining normalized frequency, although to a much milder degree.

The sharp decline of the keyword "model" is intriguing and probably more mythical. One possible conjecture is that under the paradigm of *Homo sapiens*, the pure analytical model may tend to be less useful or relevant given the explanation above; hence, it drives economists to find other ways of handling the uncertainty of a theoretical world, for example, by means of simulation, laboratory experiments, field experiments, naturally occurring experiments, or even a model-free data-driven approach. This by no means implies the extinction of the models; as a matter of fact, in some disciplines of economics, models are still very much alive, but despite this being so more space needs to be left for other accompanying approaches.[4]

[4]In a sense, this simply means that it is becoming increasingly difficult for the pure theoretical model to be accepted and published.

5.3.4 *Co-word Network Analysis*

The co-word network based on abstracts both before 2000 and after 2000 is shown in Figs. 5.6 and 5.7, respectively. The size of a word is scaled according to betweenness centrality, which is a network parameter that indicates how central a node is in a network (Borgatti 1995). We can observe that before 2000, the first tier of central words includes "model" and "equilibrium," followed by words such as "social," "uncertain," "behavior," and "optimal."

We observe a change in network structure after 2000. The co-word network shows that the centrality of "equilibrium" and "model" declined substantially. The first tier of words includes "model," "behavior," and "social," followed by "optimal," "equilibrium," and "rational." It is another way to show that the *Homo economicus* paradigm is gradually changing to the *Homo sapiens* paradigm.

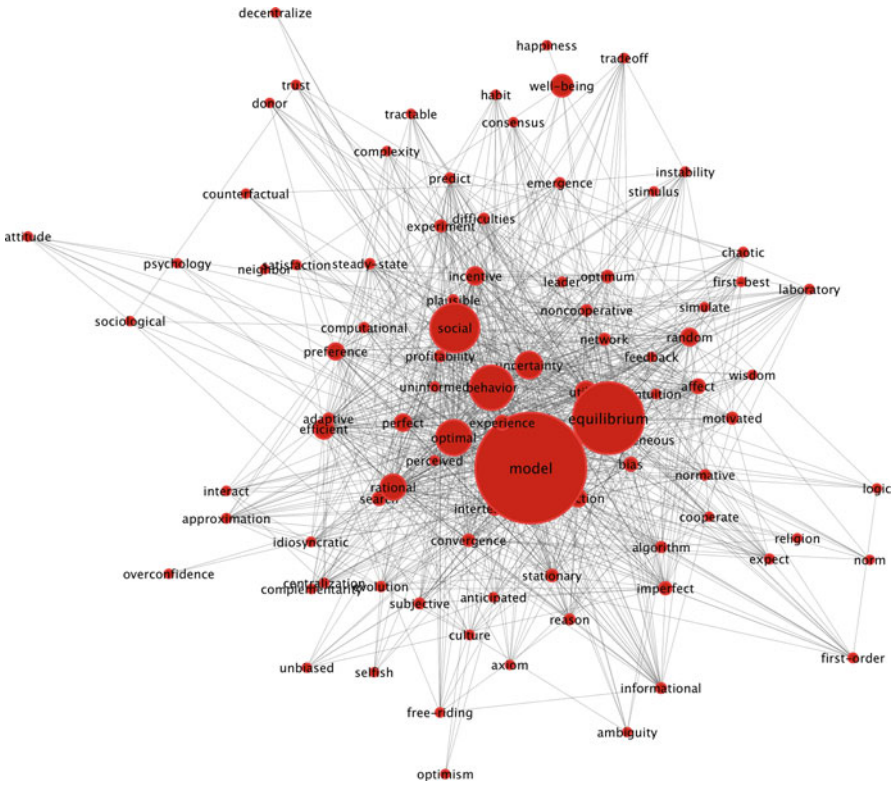

Fig. 5.6 Co-word network of economic corpus: 1992 to 1999

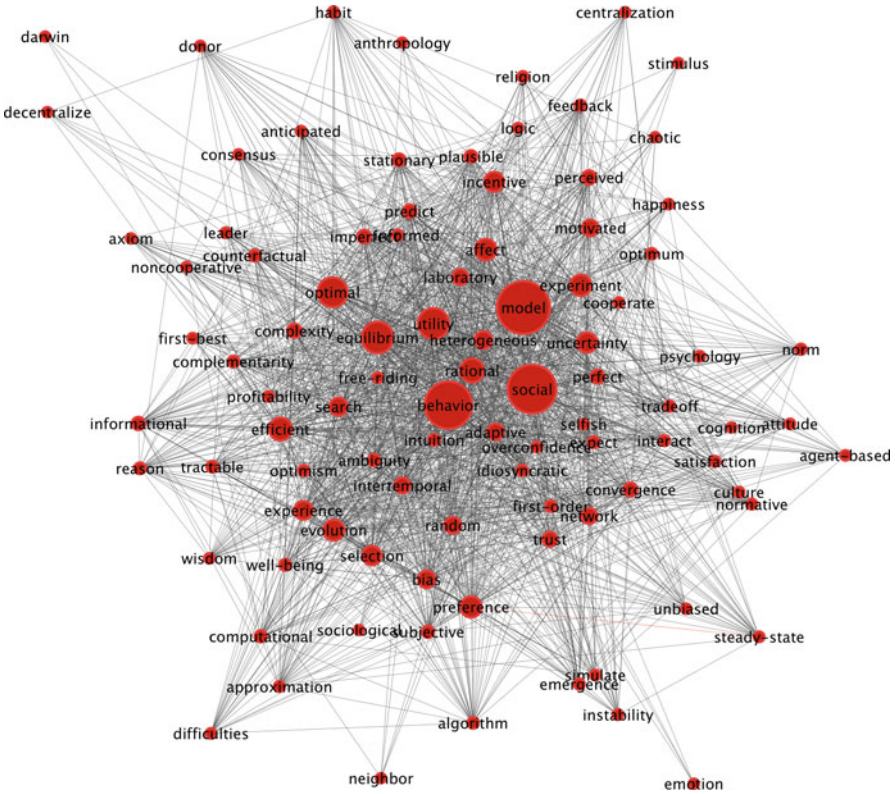

Fig. 5.7 Co-word network of economics corpus: 2001 to 2014

5.4 Concluding Remarks

In this chapter, we have applied the corpus linguistic approach to address an issue related to the paradigm shift. The historian and philosopher of science Kuhn (1962) invented the term "paradigm shift" to describe the change in the set of core concepts or the set of operating routines belonging to a scientific discipline. This term is not restricted to the sciences and was introduced to the social sciences when a similar change was experienced. While a paradigm shift can be clearly recognized after the end of the shift, recognizing it at the initial or the middle stage has not been that straightforward. For quite some time, the document-based approach has already been applied to study the process of the shift, again usually much later than when the shift settled. Nowadays, with the advances in ICT and various automated information extraction algorithms, machines can be assigned to this laborious work. Using this modern technology, we can detect or predict a shift through the constant online monitoring of the texts, posts, and human conversations in social media networks. In this way, we can know the time, the scale, and the milestones of a

paradigm shift. Hence, we can study the paradigm shift not as a subject of history, but as its contemporaneity. Our study of the paradigm shift from *Homo economicus* to *Homo sapiens*, which is still ongoing, provides a concrete illustration of this promising research methodology.

From an economics viewpoint, we can consider that keywords (symbols) are both competing and cooperating with each other to gain limited human attention. In this regard, the paradigm shift is equivalent to the shift in human attention. The keywords are, on the one hand, competing as if they want to draw a certain share of human attention. On the other hand, to achieve the above purpose, cooperating with other keywords and forming a co-word network is also important. With this conceptual framework, in this chapter, we have presented both the normalized frequency (market share) and the co-word network of keywords. The former allows us to distinguish momentum-gaining keywords from momentum-losing keywords, and the latter allows us to know the "major players" or "hubs" of a "syndicate." These two analyses together essentially provide technical characterizations of a "paradigm," which not only governs the use of our attention resource, but also dictates the organization of the attention modularized by the keywords. Hence, under this framework, a paradigm shift not only means that we pay attention to different things, but we also consider how these different kinds of attention are organized together. In this chapter, we have been able to find both characterizations, hence suggesting that a paradigm shift has occurred from *Homo economicus* to *Homo sapiens*.

There is a fundamental limitation which we do not intend to leave unnoticed. As we have mentioned before, our approach based on machines should at best read as an "assistant's job," which does not intend to replace the role that a historian can play. In fact, the foundation of this study is identifying keywords and their classifications. A machine can help us perform the first task very powerfully, but for the second one there is simply no theoretical justification on which we can rely. To some extent, this latter task is still very subjectively performed in this part, and can be problematic. Therefore, we should keep the question, regarding the soundness of the two sets of keywords established in this study, open.

Despite this possible limitation, what is found in this paper is generally insightful. On the one hand, we see the declining tendency of keywords such as "rationality," "equilibrium," and "optimal". Can this evidence alone be a sign for the decay of *Homo economicus*? On the other hand, we also see the increasing tendency of keywords such as "psychological," "emotion," "cognitive," "culture," "social," and "heterogeneity". Can this evidence alone confirm our feeling that economics has become increasingly pluralistic and has no longer carried her crown of "economic imperialism"? We do not have a definite answer, but our evidence prompts us to throw the questions out, and we believe that when digital social sciences or humanities becomes more advanced we may one day have the answer too, of course, under the joint efforts with humans.

Acknowledgements The second author is grateful for the research support in the form of *Ministry of Science and Technology (MOST) Grants*, MOST 106-2410-H-004-006-MY2.

References

Arthur, W. B. (2014). *Complexity and the economy*. Oxford: Oxford University Press.

Bianchi, M., Boyle, M., & Hollingsworth, D. (1999). A comparison of methods for trend estimation. *Applied Economics Letters, 6*(2), 103–109.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.

Blaug, M. (2001). No history of ideas, please, we're economists. *Journal of Economic Perspectives, 15*(1), 145–164.

Borgatti, S. P. (1995). Centrality and AIDS. *Connections, 18*(1), 112–114.

De Bellis, N. (2009). *Bibliometrics and citation analysis: From the science citation index to cybermetrics*. Metuchen, NJ: Scarecrow Press.

Diesner, J. (2014). ConText: Software for the integrated analysis of text data and network data. In *Social and semantic networks in communication research*. Seattle, WA: Conference of the International Communication Association (ICA).

Dohmen, T. (2014). Behavioral labor economics: Advances and future directions. *Labour Economics, 30*, 71–85.

Flowerdew, L. (2015). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes, 20*, 58–68.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*(4), 25–42.

Frey, B. S. (1997). *Not just for the money*. London: Edward Elgar Publisher.

He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends, 48*(1), 133.

Kalaitzidakis, P., Mamuneas, T. P., & Stengos, T. (2011). An updated ranking of academic journals in economics. *Canadian Journal of Economics/Revue canadienne d'économique, 44*(4), 1525–1538.

Knight, D., Walsh, S., & Papagiannidis, S. (2015). I'm having a Spring Clear Out: A corpus-based analysis of e-transactional discourse. *Applied Linguistics, 38*(2), 234–257. https://doi.org/10.1093/applin/amv019.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Lovejoy, A. O. (2011). *The great chain of being: A study of the history of an idea*. Piscataway, NJ: Transaction Publishers.

McDonald, I. M. (2009). The global financial crisis and behavioural economics. *Economic Papers: A Journal of Applied Economics and Policy, 28*(3), 249–254.

Mill, J. S. (1874). On the definition of political economy, and on the method of investigation proper to it. In *Essays on some unsettled questions of political economy* (2nd ed.). London: Longmans, Green, Reader & Dyer.

Nowak, M., & Highfield, R. (2011). *Super cooperators: Altruism, evolution, and why we need each other to succeed*. New York: Simon and Schuster.

Pumfrey, S., Rayson, P., & Mariani, J. (2012). Experiments in 17th century English: Manual versus automatic conceptual history. *Literary and Linguistic Computing, 27*(4), 395–408.

Schreibman, S., Siemens, R., & Unsworth, J. (Eds.). (2008). *A companion to digital humanities*. New York: Wiley.

Scott, M. (2001). Comparing corpora and identifying key words, collocations, and frequency distributions through the WordSmith tools suite of computer programs. *Small Corpus Studies and ELT*, 47–67.

Simon, H. A. (1962). The architecture of complexity. *Proceedings of the American Philosophical Society, 106*(6), 467–482.

Thaler, R. H. (2000). From homo economicus to *Homo sapiens*. *Journal of Economic Perspectives, 14*(1), 133–141.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Boston: Addison-Wesley.