Metaphors and Diaphors in Science Communication: Mapping the Case of Stem-Cell Research

Science Communication (forthcoming)

Loet Leydesdorff [1] & Iina Hellsten [2]

Abstract

"Stem-cell research" has become a subject of political discussion in recent years because of its social and ethical implications. The intellectual research program, however, has a history of several decades. Therapeutic applications and patents on the basis of stem-cell research became available during the 1990s. Currently, the main applications of stem-cell research are found in marrow transplantation (e.g., for the treatment of leukemia). In this study, the various meanings of the words "stem cell" are examined in these different contexts of research, applications, and policy debates. Translation mechanisms between contexts are specified and a quantitative indicator for the degree of codification is proposed.

Keywords: translation, metaphor, co-words, codification, semantics, mapping

1. Introduction

Scientists have reported significant progress in stem cell research in recent decades. The discussion about stem cells has expanded significantly beyond the scientific journals that are usually the domain of developments in science. Advances in health care promised by this line of research, together with the ethical and social implications associated with stem cell creation and exploitation in research, have attracted the attention of many groups, who, perhaps not understanding the technical literature, often use other terms to describe it. Some key terms may function metaphorically in these different contexts of use, but other terms contextualize the key terms differently so that specific meanings can be distinguished. One can expect that the construction of translation mechanisms using metaphors as messengers of meaning (Maasen & Weingart, 1995; Hellsten, 2002) is counterbalanced by other words which support the differentiation of meaning between restricted and elaborate discourses (Bernstein, 1971; Coser, 1975). In the latter case, we will below propose to use the word 'diaphors' (Weelwright, 1962; Luhmann, 1990).

One example of intense interaction between scientific and public discourses was the broader attention to stem-cell research in the address given by U.S. President George W. Bush on August 9, 2001. This was the first time an American President had gone on national TV in a special broadcast on a bioethical issue. He instructed the government-funded National Institutes of Health (NIH) to limit research funding to the 60 stem-cell lines that NIH had already recorded to

[1] Université de Lausanne, Scool of Economics (HEC) & University of Amsterdam, Amsterdam School of Communications Research (ASCoR), Kloveniersburgwal 48, 1012 CX Amsterdam, The Netherlands loet@leydesdorff.net; http://www.leydesdorff.net/

[2] Royal Netherlands Academy of Arts and Sciences, Networked research and digital information, PO Box 95110, 1090 HC Amsterdam, The Netherlands, iina.hellsten@niwi.knaw.nl

date, and ordered the Institutes not to create new lines (a process that requires using discarded human embryos). Many of the existing cell lines, however, proved unsafe for clinical trials because they had been grown on mouse media. The political decision thus interfered with the research process.

This controversy was extended further when in November 2001, President Bush convinced the U.S. Congress to ban reproductive and therapeutic cloning—a ban that would directly affect the production of stem cells. The debate on stem cells culminated in the year 2001 (Nisbet *et al.*, 2003). Wertz (2002) notes that the ban does not extend to private-sector laboratories that do not receive government funds. Only therapeutically oriented research funded by government has been banned. Stem-cell research thus provides us with a fascinating nexus of ethical, industrial, and research interests. At the interface between science and other domains in society, the words "stem cell" can be expected to have different meanings, because these different domains use different codes of communication for providing meaning to words. In particular, the degree of codification of these words is expected to vary across the domains. The sciences, for example, use more highly codified meanings than the newspapers.

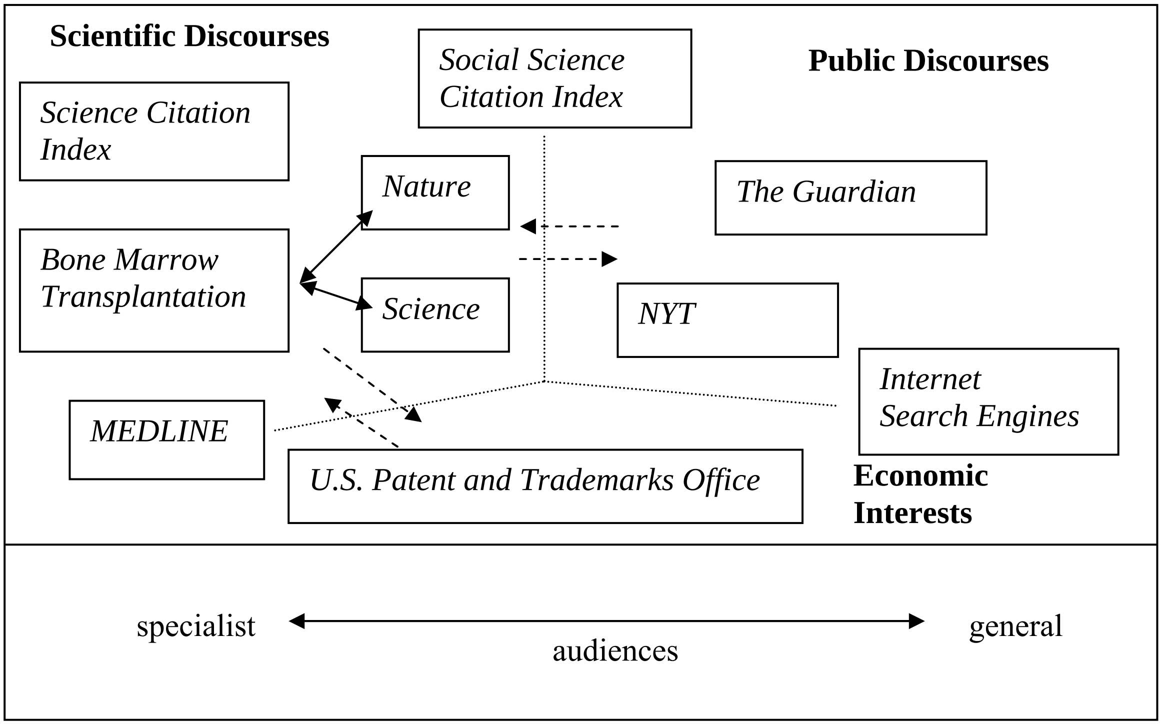

Figure 1. The different domains of research, application and policy, the related scientific discourse, economic interests and public discourses, and their expected audiences.

The aim of this article is to show in a systematic way differences in the uses of the words "stem cell(s)" in these various contexts. These differences inform us about the science communications within and between the various domains. The main focus is on three such domains: research (scientific discourse), application (economic interests), and policy (public discourse) (Figure 1).

2. Theoretical relevance

Our specific focus is on science communication, that is, the communication processes both within the sciences and between the sciences and society. Scientific communication is codified to a higher degree than non-scientific communication because it is knowledge-based (Luhmann, 1990; Leydesdorff, 1997). We are interested in the communications between the sciences and society that become public in the form of scientific publications (articles, conference proceedings, books, chapters in books, etc.), in the public representations of science in the mass media and in the Internet, and in the possible economic applications of the scientific research as it becomes public in patents. How are these contexts codified differently, and how is meaning translated by communication among domains?

Science communication is often understood as the communication of science, for example, by science journalists mediating between scientists and other audiences. Particularly when studying controversies, one can show that science is not homogenous. Scientific discourses can be deconstructed and analyzed. The focus is then on differences and changes, and therefore on variations. Our focus in this study, however, is on selection from variation, i.e., on how the information is codified in the different domains. From the perspective of communication theory (Luhmann, 1984/1995; Leydesdorff, 2001a) these processes of providing meaning to communication can be considered as processes of codification.

Codification processes can be expected to vary among social domains (Law & Lodge, 1984) and thus asymmetries are generated in terms of what words and co-occurrences of words mean in different contexts. Scientific communications are organized differently from communications in other domains (Hesse, 1980). For example, knowledge claims in science are controlled via a process of (possibly anonymous) peer review. Gilbert and Mulkay (1984) showed that the perceived status of a communication in science changes during these processes of evaluation. Scientists tend to use an empirical repertoire to explain errors and a rationalist repertoire for explaining success, but from a hindsight perspective. An analysis of scientific discourse and its specific dynamics is therefore needed (Mulkay *et al*., 1983).

For example, the knowledge claim contained in a patent is certified by a process of examination to the extent that the patent can be litigated in court (Granstrand, 1999). Within science, internal criteria of validity are often more important than external ones (Biagoli & Gallison, 2003). Scientific references in patents and in scientific articles can be expected to have different functions and meanings accordingly (Bhattacharya et al., 2003; Grupp & Schmoch, 1999; Leydesdorff, 2004). At the interface between science and society one can additionally expect an ongoing process of de-differentiation or contextualization when different types of knowledge claims have to be recombined in the process of shaping public opinion and political decision-making (Gieryn, 1999; Nowotny *et al*., 2001). The political process, however, contains also its own codification (Guston, 2000; Jasanoff, 1990), and different media can be expected to vary in terms of what can be mediated (Luhmann, 1996/2000; Leydesdorff, 1993; McLuhan *et al*., 1969).

Although the use of a common language assumes that one can translate from one context into another (Habermas, 1981), this integration is empirically traded off against differentiation. Differentiation is needed for processing the complexity reflexively. However, differentiation can only be reproduced if the differences are also codified. The system and its subsystems thus translates recursively and continuously by using words for the mediation and by providing these words with different meanings at interfaces.

3. Indicators of meaning

Callon *et al.* (1983) proposed using words and co-occurrences of words for mapping empirically the translations in the dynamics of science, technology, and society. In the 'sociology of translation' (Callon *et al.*, 1986; Law & Lodge, 1984) co-occurrences of words (co-words) have been considered as the carriers of meaning across different domains. Words, however, are ambiguous as some words refer to several objects and sometimes there are several terms for one object. Languages contain both polysemous and homonymous words. Co-words are distributed and, therefore, contain an uncertainty. Furthermore, new techno-scientific developments can be expected to change contexts innovatively, and the words may change in meaning accordingly.

Words are contained within sentences that provide them with meaning (Bar-Hillel, 1955; Hesse, 1980; Leydesdorff, 1995, 1997). Specific keywords, however, may function as carriers of meanings between science and society (Hesse, 1988; Maasen & Weingart, 1995). This process remains highly mediated, for example, in terms of patents, while the sciences themselves are intellectually organized in specialties that tend to maintain strong boundaries among them (Kuhn, 1984; Whitley, 1984). The communication of specialized knowledge on both sides of an interface may require a shared representation (e.g., in terms of words), but one expects this representation to be integrated differently on either side.

We have distinguished (Hellsten & Leydesdorff, 2004) between two types of carriers of meaning as reflexive mechanisms for the comparison of co-words across the domains: metaphors and diaphors. On the one hand, metaphors can be considered as functioning symbolically, i.e., as 'reflexive actants' in the network of words, and on the other hand, sub-symbolic distributions of words, diaphors, function as interactions among the different (sub)centra in the network. The translation in science communication may, therefore, function both symbolically (metaphors) and sub-symbolically (diaphors). We suggest that 'metaphors' and 'diaphors' can be considered as tools of intermediation that channel meanings across different arenas in the communication of science because they both contribute to the carrying of a set of relations from one domain to another.

Some key terms, such as "stem cells" may function metaphorically in one context and diaphorically in another context. While a metaphor can be used to make the translation by focusing on a similarity, the diaphor highlights a difference (Weelwright, 1962). Luhmann (1990) used the word 'diaphor' to distinguish analytically between words that carry meaning and words that contribute to boundary construction between domains of communication. Thus, the question of translation can be made an empirical one about the extent to which words and co-words can be used to indicate differentiation and integration in the communication.

The increased availability of online resources makes it possible nowadays to map all the relevant domains in terms of co-occurrences and co-absences of words. Mappings of co-occurring words can show the various contexts of codification over time using a defined context (Small and Greenlee, 1986; Small, 1999) or across contexts at a specific moment in time. Here, we focus on different contexts at a specific moment of time. Elsewhere (Hellsten, 2003; Hellsten & Leydesdorff, 2004) we have mapped co-words also over time. The networks are constructed in terms of *relations* among words, but the words are also *positioned* in the maps. The maps can thus function as a representation of semantic fields. These next-order structures can be compared with one another in terms of the degree of codification.

4. "Stem cells" as a topic in different discourses

The words "stem cell(s)" can be retrieved abundantly in, for example, the *MEDLINE* database at <http://www.pubmed.gov/> with 126 hits during the 1960s,[3] 890 hits during the 1970s, and 1,648 hits during the 1980s. The research on this topic began to grow spectacularly during the 1990s (5,939 hits). The growth of the field in terms of the retrieval for these search terms as title words in the *Science Citation Index (SCI)* is even more spectacular with more than 2,000 documents in the year 2000, and more than 3,000 in the year 2002.[4] These growth patterns fit exponential curves ($r^2 > 0.95$; the curves are not shown here).

Patents with "stem cell" in the title did not appear before the 1990s and did not gain momentum before 1997 (with 25 patents in this single year retrievable from the database of the U.S. Patent and Trade Office at <http://www.uspto.gov/ >). The reflection in the social sciences and the humanities followed this development at an even later date. As noted, the public debate about "stem cells" flourished in 2001, when the U.S. government restricted research to a number of cell lines. In Figure 2, this third dimension—i.e., the public debate—is illustrated by using the number of retrievals in the *New York Times* as an additional indicator.

[3] The first publication retrieved in this database with the words "stem cell" in the title is: Gurney, C.W. 1963. 'Effect of radiation on the mouse stem cell compartment in vivo.' *Perspect. Biol. Med.* 6(2):233-245.

[4] For reasons to be explained in the next section, we have used the occurrence of the string ("stem cell" OR "stem cells" OR "stem-cell") in the titles of documents.

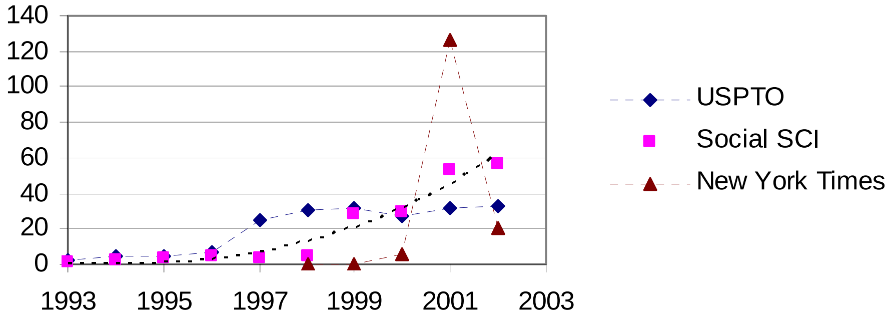

Figure 2. The industrial, social, and political relevance of stem-cell research, using various online sources.

Unlike these specific databases, the Internet provides us with an overarching representation of stem-cell research that can be accessed using search engines and meta-crawlers. Despite its well-known shortcomings (Bar-Ilan, 2001; Rousseau, 1999; Thelwall, 2001), we have used the *AltaVista Advanced Search Engine* for this study because this search engine enabled us to organize numbers of hits in terms of calendar years. Note that this representation is based on the hindsight perspective of an ongoing reconstruction, while the other databases are made permanent at the end of each calendar year (Leydesdorff, 2001b; Wouters *et al*., 2004). The *AltaVista* search engine classifies a page dynamically, that is, at the date that it was last updated.[5]

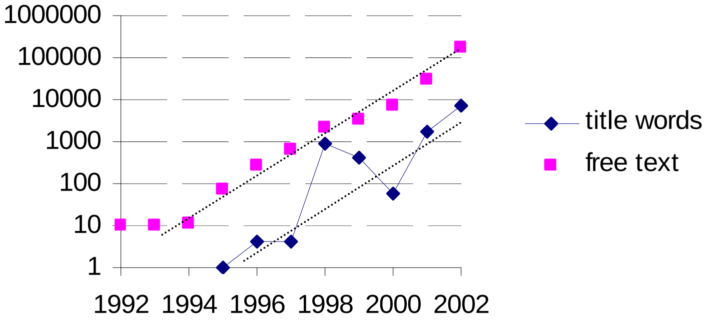

Figure 3. Number of records on "stem cell" retrieved using the *AltaVista Advanced Search Engine* on March 14, 2003.

Figure 3 shows that attention to stem-cell research has been steadily on the rise with the growth of the Internet during the 1990s. (Note that the scale of the vertical coordinate is logarithmic in this case.) The number of pages using "stem cell*" among its *title words*, however, shows a spectacular increase in 1998. This upsets the normalized time curve with more than an order of magnitude. The effect is consistent with the increased social relevance of this research as manifested in the social science literature and in the patent database (Figure 2). Public interest manifested at the Internet did not disappear in 2002 as the political debate subsided, but returned

[5] The expanded edition of the *Science Citation Index* is sometimes updated, but this was not relevant for the relatively short period of collecting data for this study.

to the normal growth pattern (which is partly an effect of the growth of the Internet and the *AltaVista* domain within it).[6]

In summary, the political intervention of the Bush administration in 2001 is most visible in the newspapers, but it is preceded by an increased visibility of the topic in both patent databases and the *Social Science Citation Index* as reflections of the social (and economic) relevance of this research. As recently as 29 September 1997 *The Scientist* opened with a lead article complaining that the announcement by a team at Johns Hopkins that they had cultured human embryonic stem cells, had received 'surprisingly little notice in the media' (Lewis, 1997). The media attention in the newspapers thus reflects the political agenda more than the scientific, social, and economic interest of developments, while the Internet search engine reflects the latter as well.

We shall proceed with mapping the meaning of "stem-cell" (OR "stem cell" OR "stem cells") as retrieval terms for titles in the various domains distinguished above. The words "stem cell" are expected to function differently in the three domains of analysis: in the U.S. Trade and Patent Office database "stem cells" are expected to refer to applications with potentially high economic values (Granstrand, 1999). In the scientific domain, "stem cells" are expected to refer to the research programs, while the policy domain can be expected to focus on the ethical and political implications of stem-cell research and its possible future applications. This is reflected in the newspaper coverage.

5. Methods and materials

The systematic comparison is pursued for the calendar year 2001. Preliminary research has confirmed our expectation that only title words are specific enough for tracing the topical concept precisely (Leydesdorff, 1989). Abstracts often mention "stem cells" as one area of application among others. Thus, the recall would be less precise if abstract words were included. The study is structured in three parts, using: (a) the patent data as a representation of applications with potentially economic value, (b) the citation index databases as a representation of the research domains and the scholarly discourse, and (c) internet data, including online newspaper archives as a representation of the policy domain (Figure 1 above).

Patent data are brought online by the U.S. Patent and Trade Office (USPTO) and by the European Patent Office (EPO). The latter database also contains the data of the World Intellectual Property Organization. However, the European and world patents are not fully standardized and partly in other formats, while the U.S. database is standardized, organized in hypertext mark-up language, and accessible for searching by robots.[7] Furthermore, the U.S.

[6] The *AltaVista* search engine was restructured in 1999. Thelwall (2001) reports that the stability of the retrieval was improved after this restructuring. Recently (April, 2004), *AltaVista*'s search engine was merged into the *Yahoo!* Search Engine.

[7] The USPTO states the following limitation at http://www.uspto.gov/patft/help/notices.htm: "These databases are intended for use by the general public. Due to limitations of equipment and bandwidth, they are not intended to be a source for bulk downloads of USPTO data. Bulk data may be purchased from USPTO at cost (see the USPTO Products and Services Catalog). Individuals, companies, IP addresses, or blocks of IP addresses who, in effect, deny service to the general public by generating unusually high numbers (1000 or more) of daily database accesses (searches, pages, or hits), whether generated manually or in an automated fashion, may be denied access to these

database is often used in scientometric research because it standardizes the presence of other nations in a single representation (Jaffe & Trajtenberg, 2002; Narin & Olivastro, 1988). This database allows, among other things, for the retrieval of citation patterns in terms of both the previous patents cited and the scientific (that is, non-patent) literature cited.[8]

As other sources of data we used the online versions of the *Science Citation Index (SCI)* and the *Social Science Citation Index* (*SSCI*), the *MEDLINE* database of the U.S. National Library of Medicine (at http://www.pubmed.gov), and the online databases of journals and newspapers at their respective sites. Where necessary specific routines were written for the retrieval, parsing, and database management. As noted, we used the Advanced Search Engine of *AltaVista* for mapping the Internet because this search engine enables us to combine Boolean operators with time delineations so that time-series can be constructed, albeit from a hindsight perspective (Leydesdorff, 2001b; Wouters *et al.,* 2004). One well-known disadvantage of this search engine is its potential instability over time (Rousseau, 1999; Thelwall, 2001). Therefore, the search was repeated several times during two consecutive days and a moment was selected when the results could be replicated.

Throughout this study, various programs were used for the database management. Graphic representations are based on Pajek and UCINET 6.[9] SPSS was used for the statistical analyses. The visualization guides the exploration in three steps:

1. the networks of co-occurrences of words can be visualized before normalization. This provides us with a representation of the variation and observable structures;
2. the vectors of the word distributions are related using the cosine for the normalization (Ahlgren et al., 2003; Salton & McGill, 1983).[10] The vector-space model enables us to show the clustering among the words;
3. the matrices of title words versus documents are factor analyzed using Varimax rotation and forcing six factors.[11] The factor structure will enable us to quantify the degree of codification in the network.

The mappings are optimized for the visualization using pragmatic cut-off levels of word frequencies in order to keep the maps readable (using approximately 120 words as the maximum). The visualizations are based on using the algorithm of Kamada & Kawai (1989) as it is available in Pajek.[12]

servers without notice."

[8] For reasons of consistency, the stopword list available at http://www.uspto.gov/patft/help/stopword.htm was used throughout this study as a standard corrective to the inclusion and exclusion of common words. Otherwise, the words are corrected only for the plural 's.'

[9] The homepage of Pajek can be found at http://vlado.fmf.uni-lj.si/pub/networks/pajek/

[10] Salton's cosine is defined as the cosine of the angle enclosed between two vectors *x* and *y* as follows:

$$\text{Cosine}(x,y) = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2)*(\sum_{i=1}^{n} y_i^2)}}$$

[11] The choice of six factors is heuristic in this stage, but the results will allow us to make comparisons between factor solutions more systematically in a later section.

[12] This algorithm represents the network as a system of springs with relaxed lengths proportional to the edge length.

6. Results

6.1 Patents

In the database of the U.S. Patent and Trade Organization (USPTO), the words “stem cell” were used as title-words in 32 records for 2001. These patents contained 98 unique title words, of which 67 occur only once. These single occurrences were discarded,[13] as were the original retrieval words “stem” and “cell” (or “cells”) because these two retrieval terms would tie the whole network by definition into a single cluster. This left us with 29 unique title words for the further analysis.

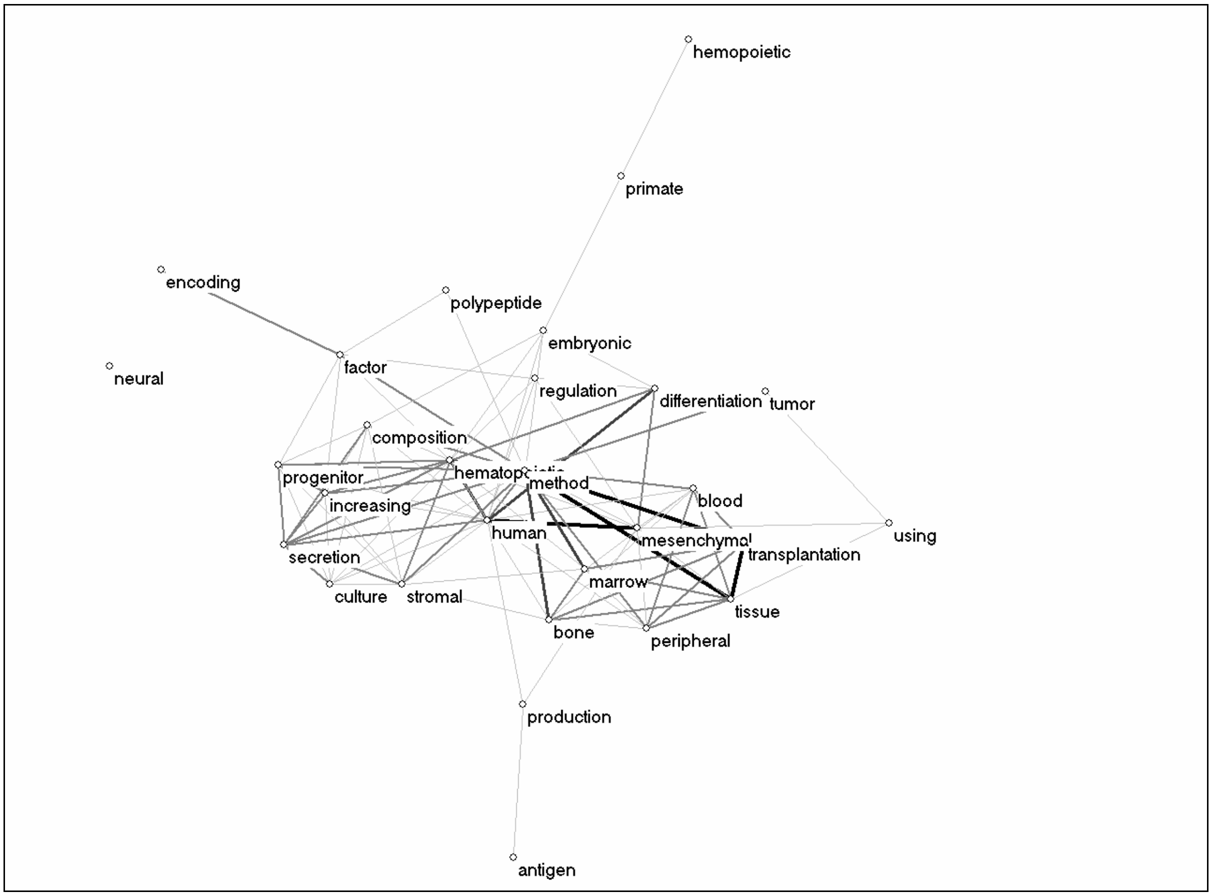

Figure 4. Co-occurrence map of 29 patent words in 2001. The line thickness is proportional to the number of co-occurrences between the words.

Nodes are iteratively repositioned to minimize the overall ‘energy’ of the spring system using a steepest descent procedure. The procedure is analogous to some forms of non-metric multi-dimensional scaling. A disadvantage of this model is that unconnected nodes may remain randomly positioned across the visualization. Unconnected nodes are therefore not included in the visualizations below.

[13] Title words that occur only in one single document only contribute to the variation, but not to the network structure.

Figure 4 maps the co-occurrences between the title words of these documents. As can be expected, the co-occurrence map shows the most frequently occurring words as central in a star-shaped network, while it also exhibits specific clusters of words. In other words, one can expect two competing effects in the representation of observed data: hierarchical integration into star-shaped networks versus differentiation and grouping into specific clusters. In this patent set, for example, ‘method’ and ‘human’ have a central position, while ‘bone,’ ‘marrow’, ‘transplantation,’ ‘tissue,’ and ‘peripheral’ form a cluster of tightly related words. Some frequently occurring words (e.g., ‘factor’) are nevertheless specific in their pattern of relations. ‘Neural’ does not co-occur with any of the other words occurring more than once.

Normalization of the word frequencies in terms of cosines among vectors counteracts the stellar form of the network around the most frequently occurring words (Salton & McGill, 1983). In other words, the structural dimensions (as different from hierarchical relations) are more clearly visible (Burt, 1982). For example, the cluster of words like ‘blood,’ ‘marrow,’ ‘transplantation,’ etc., is more pronounced in the normalized representation (not shown here). This cluster reflects that blood and marrow transplantation, used in the treatment of several types of cancer, is currently the main application of stem-cell therapy.

The results of the factor analysis confirm that the overall structure in the word patterns corresponds largely with the visualization before (and similarly after) normalization. This indicates that a strong normalization has already taken place when the patents were provided with title words. These words were highly codified when entered into the patent database. In other words, the words are used in the context of a specialist discourse for a specific audience (e.g., the patent examiners). The factor analysis confirms this observation. Six factors explain 61.0% of the variance in the matrix, but the first three factors already explain 42.8%. This skewed distribution of the relative weight of the factors means that the word usage in this database is highly codified (Leydesdorff, 1997).

6.2 The *Science Citation Index* and *MEDLINE*

Of the 2,634 documents indicated as retrievable using our search terms in the online version of the *Science Citation Index* 2001, 2,630 titles could actually be retrieved. These titles contain 4,524 unique words, of which 155 occur in a frequency $\geq$ 30. Eleven more words were removed because they did not have co-occurrence links within the set above a frequency threshold of ten. When the network of co-occurrence links between the remaining 144 words is visualized, this leads to a densely knit network with most of the main words in the center (Figure 5). One specific grouping with words from experimental biology is visible at the left side of the map. This group is related to the main group via the word ‘human’ because of the use of human embryos.

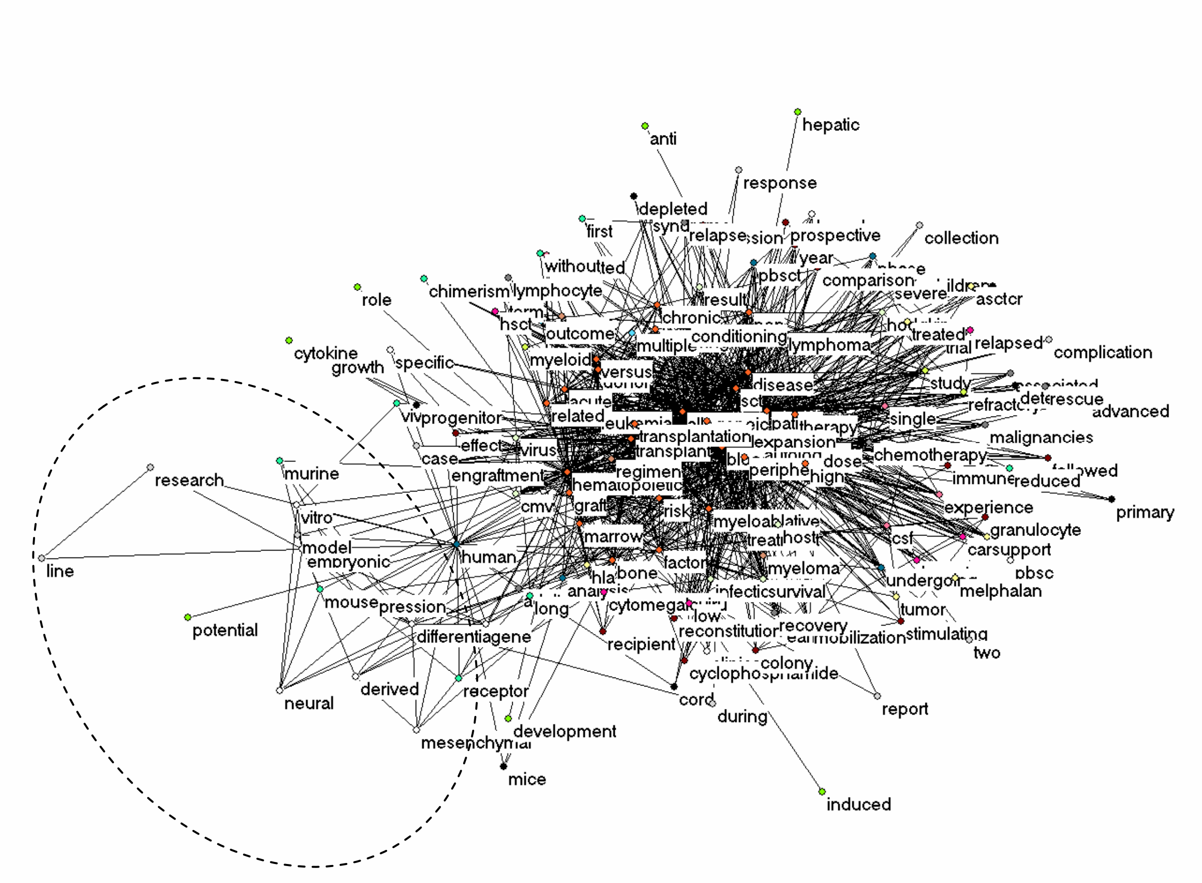

Figure 5. Co-occurrence map of 144 title words from the *Science Citation Index* with ten or more linkages within the set.

In this case normalization clarifies the underlying structure (Figure 6). By choosing a threshold at a value of the cosine ≥ 0.2, 96 words inform us about the various clusters in this set, such as a separate cluster 'cytomegalo-virus-infection,' and a central group indicating 'high-dose-chemotherapy.'

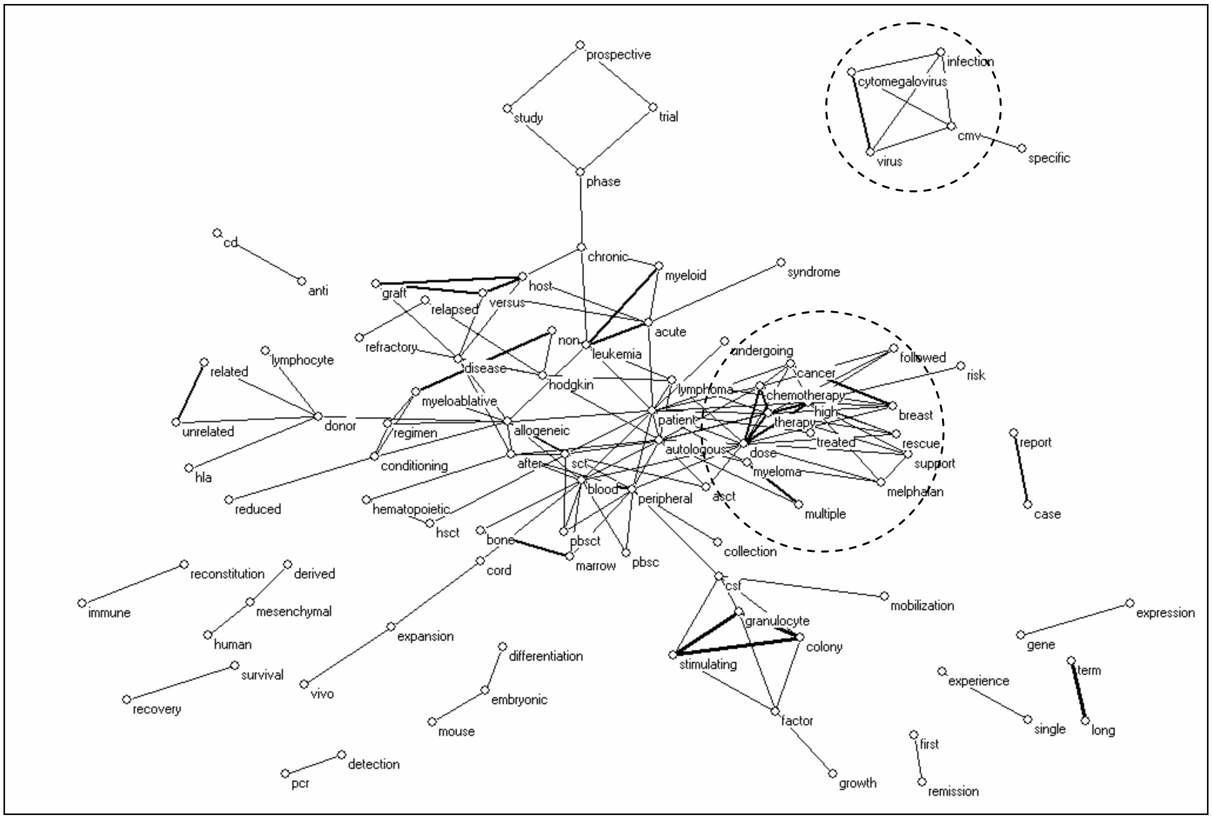

Figure 6. Normalized word distributions of 96 title words in the *Science Citation Index* with a cosine ≥ 0.2 between them.

High-dose chemotherapy, for example, is used for lymphoma cancer patients in order to destroy the bone marrow before stem cell therapy. Cancer patients receiving autologous stem cell transplants often run the risk of contracting a serious infection called cytomegalovirus infection. The various clusters of words show the different topics of research as represented in the *Science Citation Index (SCI)*.

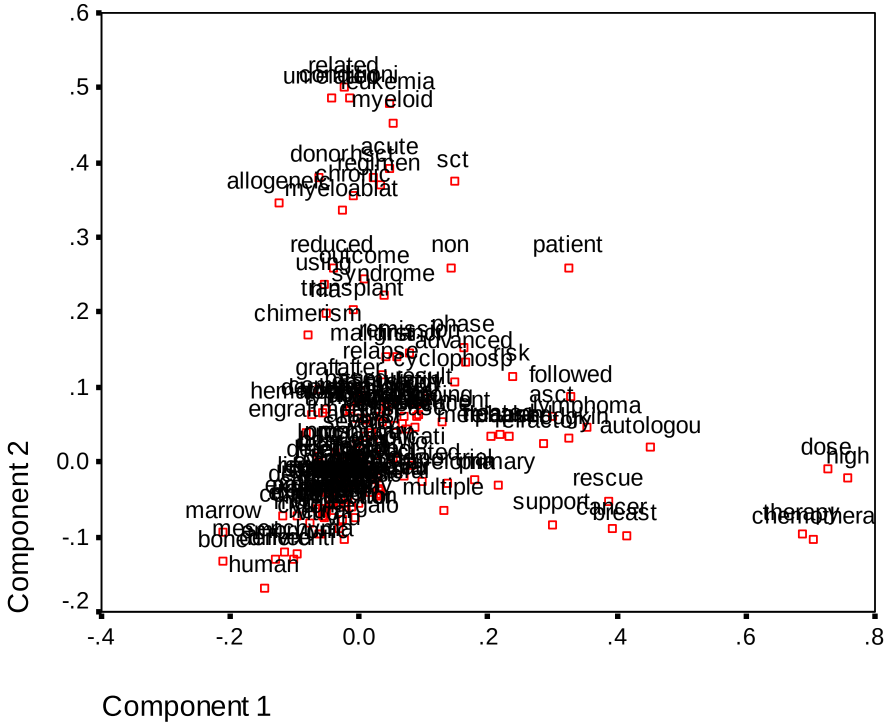

Figure 7. Factor plot for the first two factors using 144 title words from the *Science Citation Index* 2001.

Factor analysis of the word patterns in the data matrix shows that individual factors do not explain a substantial part of the variance: the first factor, for example, explains only 3.4%. This is extremely low, but the loadings of words on different factors are yet highly specific (Figure 7). For example, the “high-dose-chemotherapy” cluster is visible as the first component. However, the “cytomegalovirus-infection” loads on a fifth factor and is thus not visible in this factor plot. This is one of the advantages of using visualization tools based on network analysis. The multi-dimensional landscape is flat, but with specific peaks that correspond to the clusters visible in Figure 6.

The data collected from the *MEDLINE* database exhibits a pattern very similar to those based on the *Science Citation Index.* However, other words are clustered. This is a consequence of the different selection criteria of these two databases. Although the factorial structure is as flat in *MEDLINE* as in the *Science Citation Index*—with the first six factors explaining only 17.9% of the variance in the matrix—the specificity of the words along the axes is somewhat stronger than in the case of the *Science Citation Index*. This can be explained in terms of the higher degree of disciplinary specialization in this database. *MEDLINE* is dedicated to the specific field of bio-medicine more specifically than the *Science Citation Index* and one would therefore expect a more precise codification.

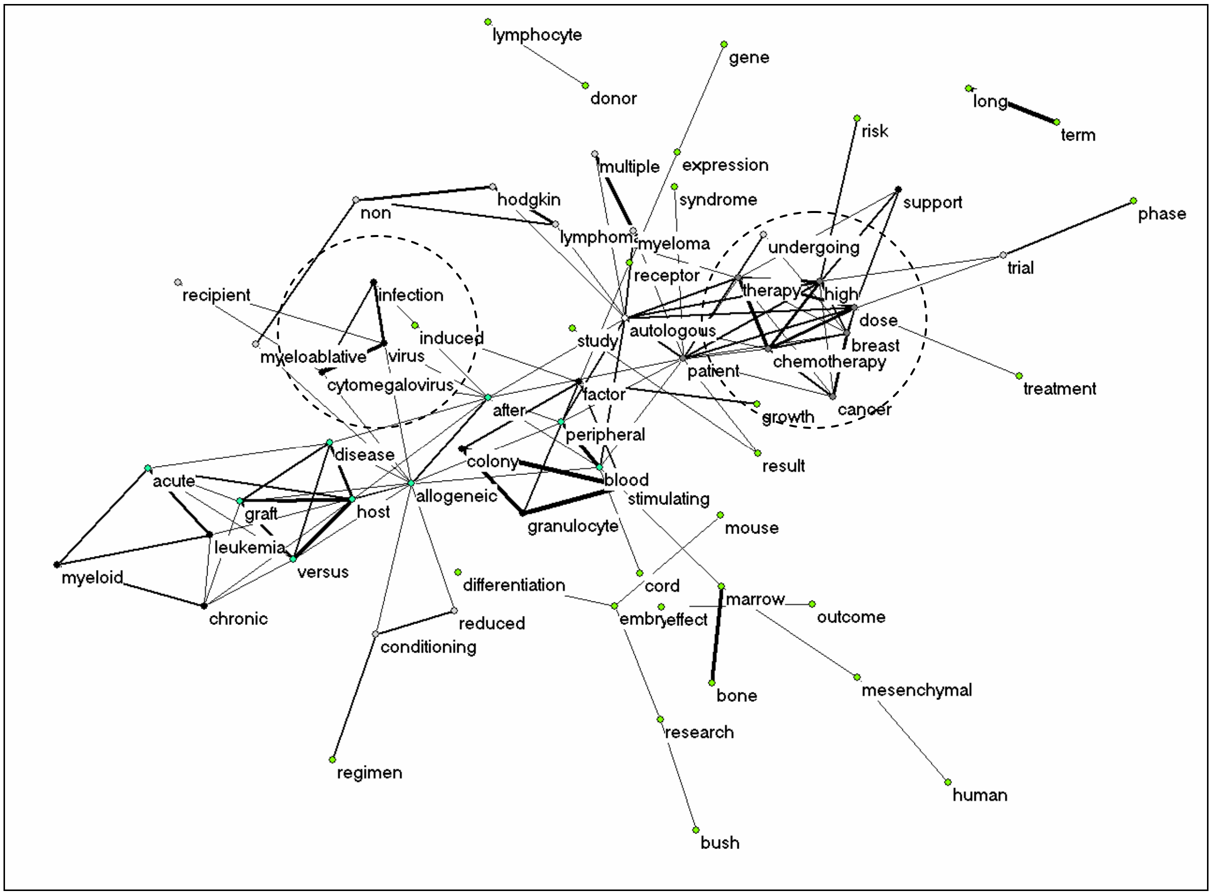

Figure 8. Representations using 67 title word distributions from *MEDLINE* 2001 with cosine ≥ 0.2.

Figure 8 shows the landscape in the *MEDLINE* database using 67 words (from 1,597 documents) which pass a threshold of 20 occurrences or more with a cosine ≥ 0.2. The two clusters which were highlighted in Figure 6 are again indicated. In summary, *MEDLINE* and the *SCI* reflect differently on the disciplinary discourses, but they are structurally rather similar in providing us with a lot of information. They can both be considered as representations of ongoing research activities. Consequently, they mainly exhibit the variation in the semantics across the relevant research fronts. The clusters represent the various research programs reporting in the journals indexed by the *Science Citation Index* and *MEDLINE*, respectively.

6.3 The *Social Science Citation Index* and the *Arts & Humanities Citation Index*

In addition to the 52 documents which could be retrieved from the *Social Science Citation Index* 2001 using our search terms, one more document was found in the *Arts & Humanities Index.* These 53 documents contained 204 unique words, of which 161 were single occurrences. Because the two retrieval terms were also not used, we pursued the analysis with a set of (204 – 161 – 2 =) 41 title words.

Somewhat surprisingly, the co-occurrences of words in this data are codified to such an extent that the mappings are not further improved by the normalization. Figure 9 shows the mapping of the network of co-occurrences normalized in terms of the cosines among the vectors. Because of the low numbers no thresholds had to be used. The factor analysis confirms a pattern very similar to the one found in the USPTO patent database above: six factors explain 57.7% of the variance.

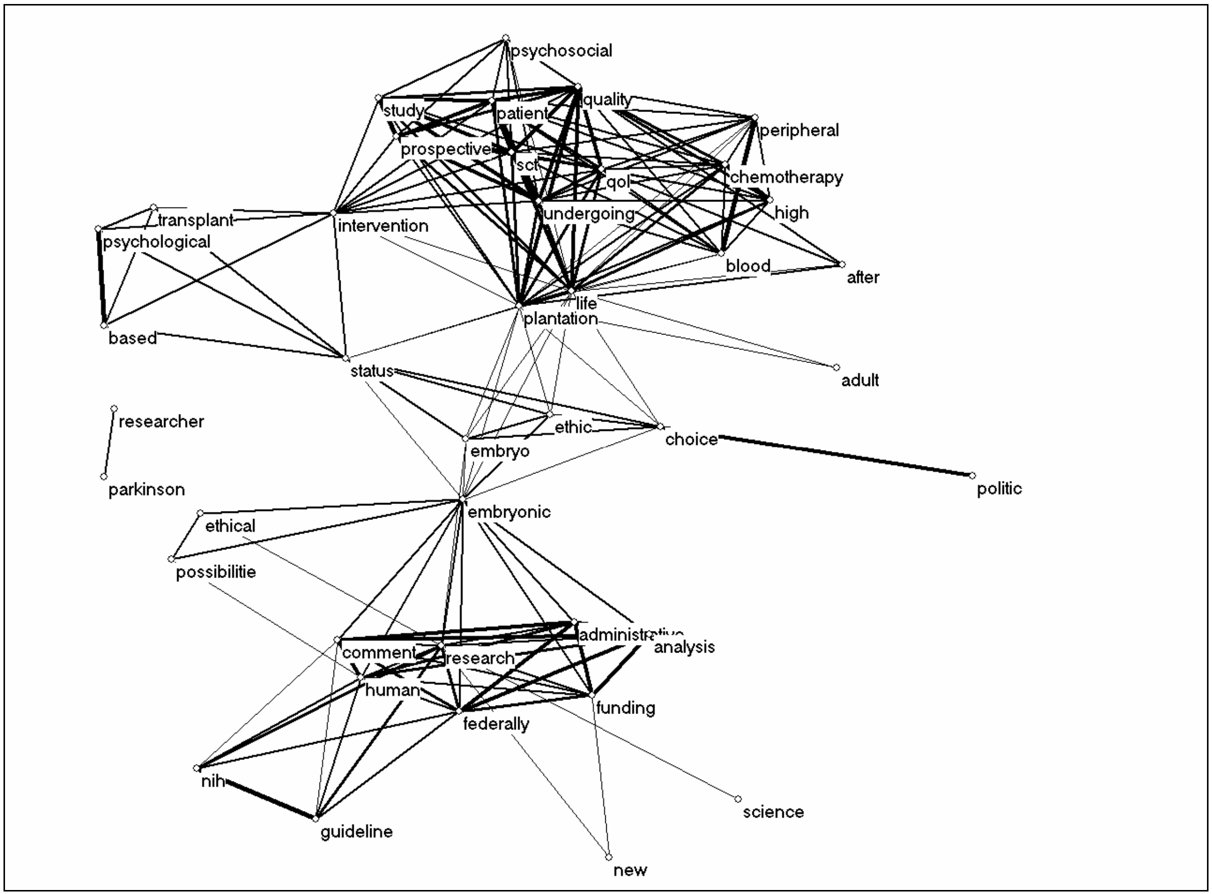

Figure 9. Cosine-based mapping of 41 words from the *Social Science Citation Index* and the *Arts & Humanities Index.*

Both pictures (before and after normalization) exhibit a large cluster that focuses on research about the quality of life of cancer patients who have undergone various forms of therapy. A smaller subset of words referring to psychological research is related to this cluster by the word 'intervention.' The other large group is focused on regulation by the federal administration of the U.S.A. In between is a group of words pointing to the ethical issues involved. In summary, the topical subjects under study are not strongly connected since the discursive reflections are organized along a number of disciplinary lines.

6.4 *The Internet and the newspapers*

6.4.1 *AltaVista*

Between March 12 and March 14, 2003 several runs with the Advanced Search Engine of *AltaVista* using our specific search terms returned variably between 1,650 and 1,900 hits for the year 2001. A run with a stated recall of 1,856 records on March 12 was used for the further analysis of the co-occurrences of title words. One thousand sixty unique titles could be harvested from this run. The web-pages range from providing information for stem cell therapy patients (e.g., at www.cancerbacup.org.uk/info/bone/bone-5.htm) to religious discussion groups interested in the ethical issues involved in the cultivation of embryonic stem cell (e.g., at www.biblelessons.com/abortion.html) and the potential future applications of stem cell therapies (e.g., at www.womens-wellness.com/wellness/199811/msg00007.html).

The retrieved documents contained 1,300 unique title words. Of these title words only 122 occurred seven or more times. As in the above cases of the *Science Citation Index* and *MEDLINE*, these title words form mainly a dense and star-shaped network of relations. When one raises the threshold, fewer words are drawn into the central core, but the format remains star-shaped. Figure 10, for example, provides a representation of the 92 words remaining when the links with values lower than five are removed. Smaller groups of tightly linked words remain visible as specific relations, but the number of isolates increases rapidly as the threshold is increased. The word ‘research’ dominates the star-shapedness of this network because of the frequent occurrence of the string “stem-cell research.” (As above, the words ‘stem’ and ‘cell’ were not included.)

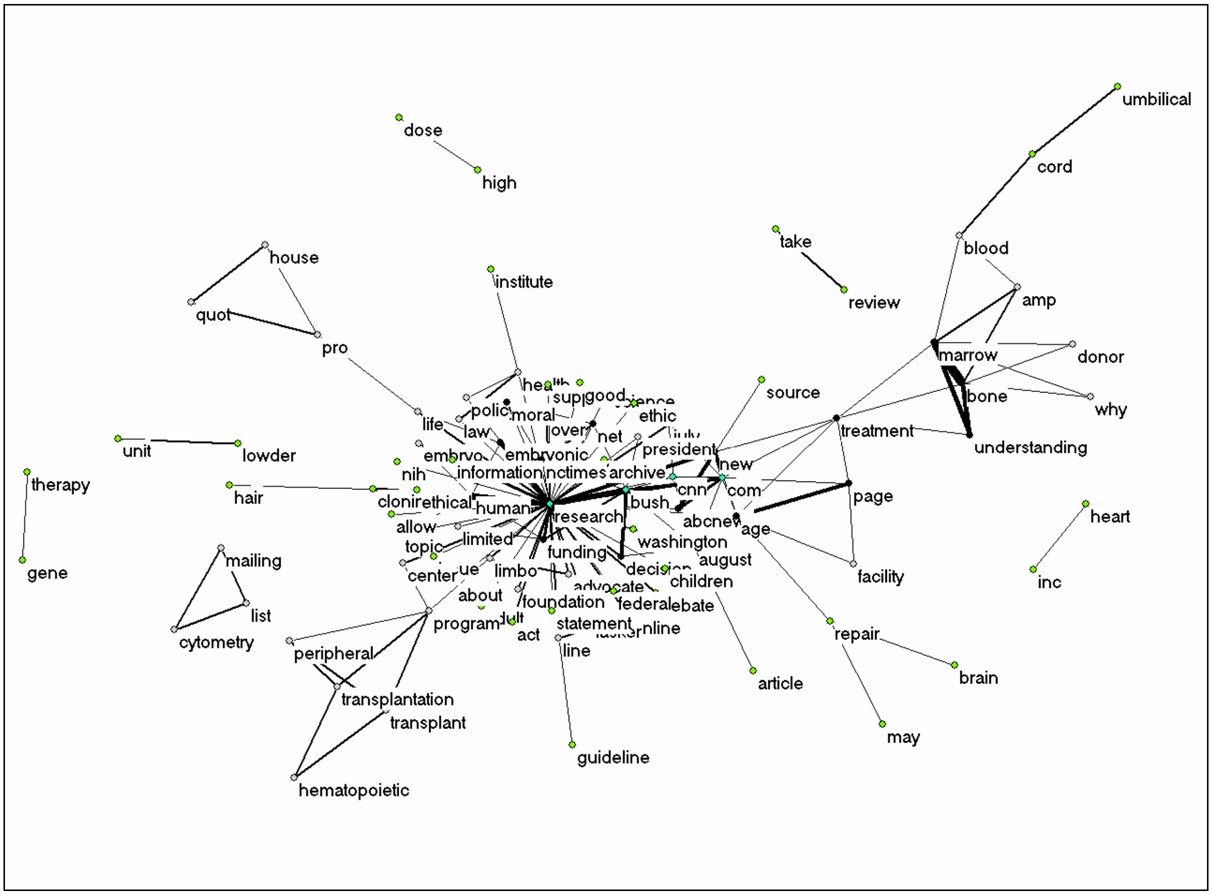

Figure 10. Co-word relations among 92 words in the domain of the *AltaVista* search engine with a frequency of five or more (following removal of the isolates).

The word 'research,' however, never relates to one of the other words above the threshold of the cosine ≥ 0.2. The normalization thus removes this size-effect completely. The normalized picture (Figure 11) exhibits the structure of related words indicating specific topics in the discussions at the Internet. The political discussion relates the various issues, but the words representing topics (e.g., why to be a donor of bone marrow) are more densely interconnected as clusters. These topics are no longer corresponding to the research topics which we have seen emerging in the previous analyses.

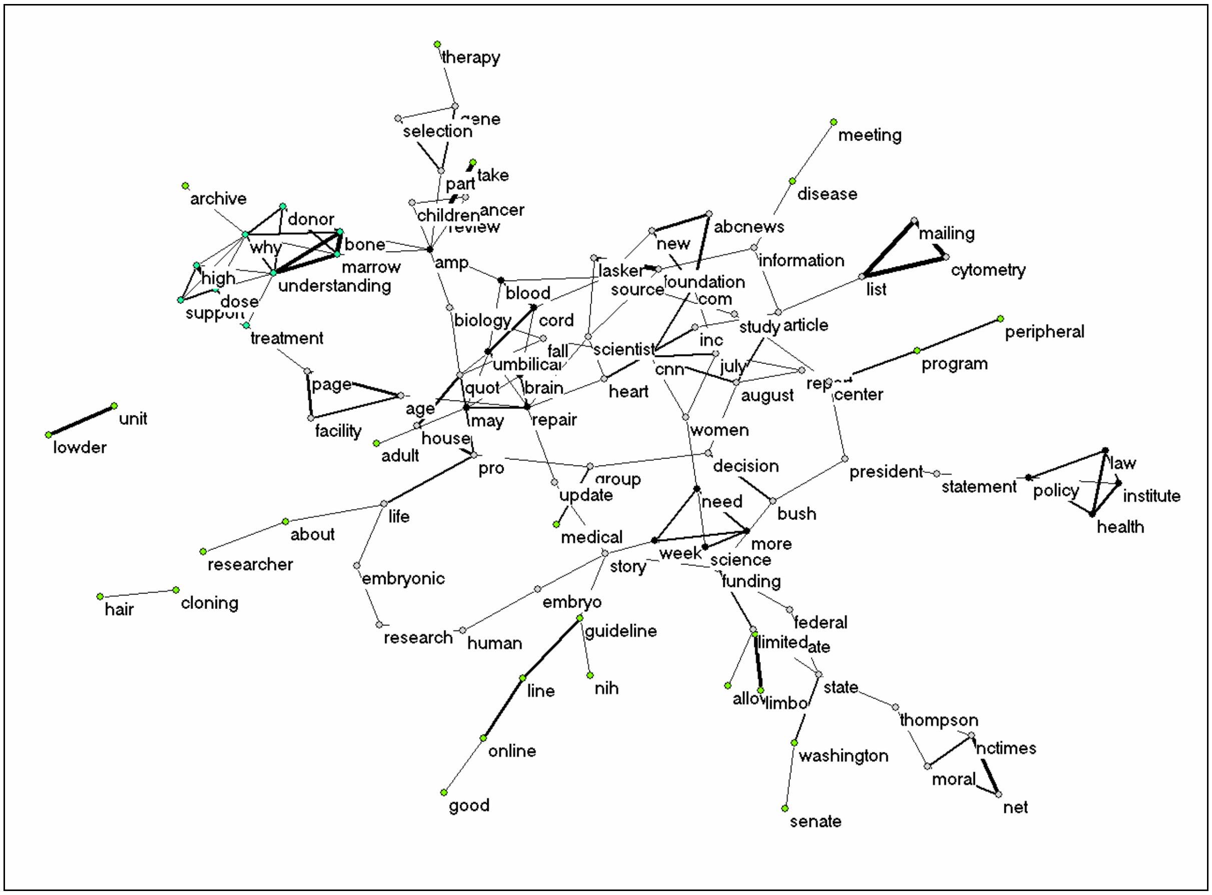

Figure 11. Hundred-nine title words from the Internet (*AltaVista*) related at a threshold level of the cosine ≥ 0.2.

Using these title words, six factors explain 13.8% of the variance. As in the previous cases, factors indicate tight co-occurrences of specific words. Thus, we observe a landscape that is different from the representation using the *Science Citation Index* or *MEDLINE* only in that relations between the clusters are apparent. In other words, the flat landscape contains mainly small, but narrow rifts instead of relatively isolated hills. Let us now turn our attention to the uses of the words "stem cell(s)" in the policy domain, and the related public discourse.

6.4.2 The New York Times

In the online archive of the *New York Times* (at http://query.nytimes.com/search/advanced?srchst=nyt), the words "stem cell(s)" were mainly used in articles and news items about President Bush's ban on stem-cell research and ethical issues related to research. The item could be retrieved in 127 headlines in 2001.[14] These titles contain 274 unique words, of which 191 were single occurrences. After deletion of 'stem' and 'cell,' 81 title words entered into the analysis.

[14] The retrieval was 89 for "stem cell" and 38 for "stem cells".

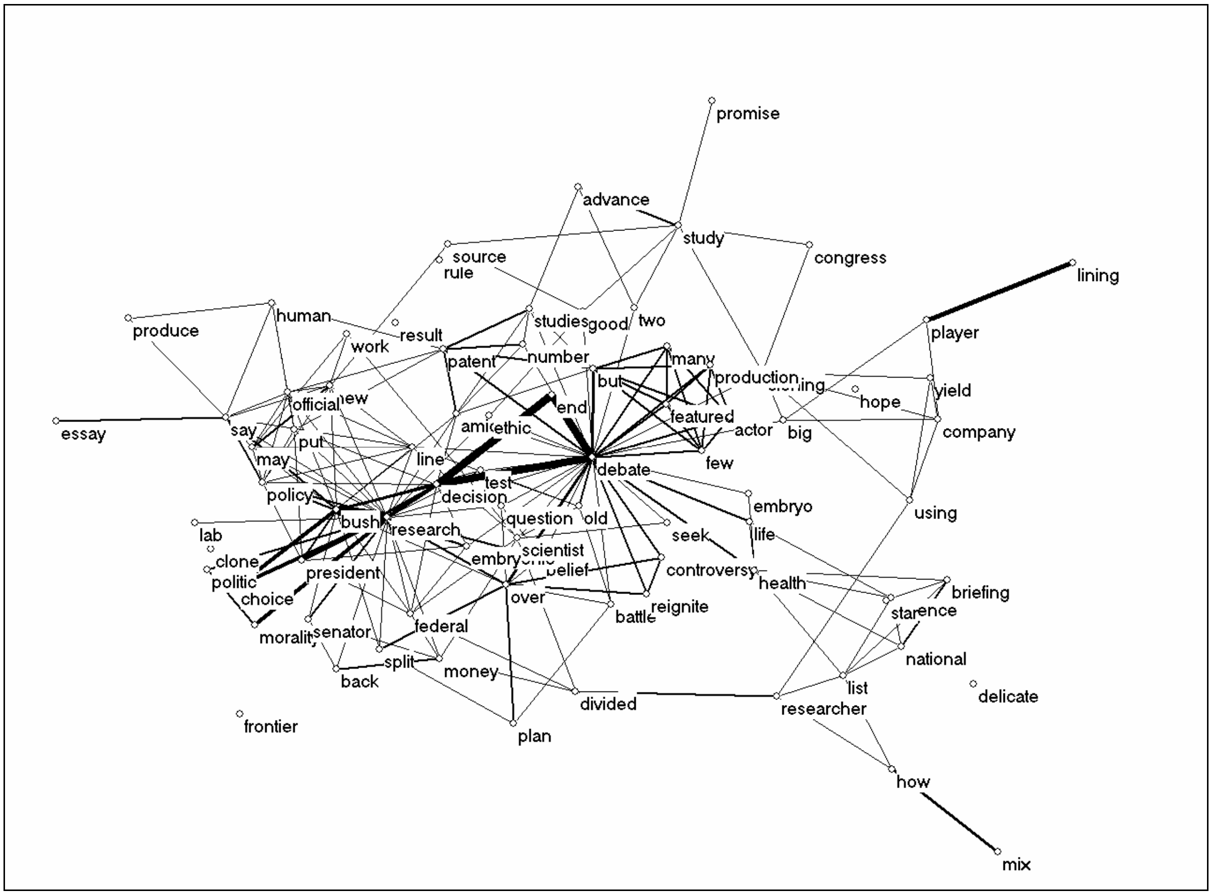

Figure 12. Cosine-based map of 81 title words from the *New York Times* in 2001 (no threshold applied).

The co-occurrence map exhibits a star-shaped networked around the word 'debate.' This star shape is moderated when the word distributions are normalized in terms of the cosine (Figure 12). The clusters of meaningful words are then more visible, but the effects of the normalization are limited. Six factors explain 27.5% of the variance. Thus, the codification is less strong than in the *Social Science Citation Index* or the *USPTO* database, but considerably higher than in the databases of the *Science Citation Index* and *MEDLINE*. Before we draw conclusions from these results, let us check another newspaper. We chose *The Guardian* because this British newspaper is freely accessible in the online version.

6.4.3 The Guardian

When "stem cell" or "stem cells" were used for searching the archive of *The Guardian* (at http://www.guardian.co.uk/Archive/0,4271,,00.html) for the year 2001, 167 articles could be retrieved. The titles in this newspaper are shorter and try to catch the readers' immediate attention. These articles also deal with a wider spectrum of issues than the articles in the dataset for *New York Times.* Consequently, more unique words are used: 322 of the 421 unique words are single occurrences. Another two words are not related at the network level. Above the threshold of cosine ≥ 0.2, 91 words provided us with the materials for the following visualization.

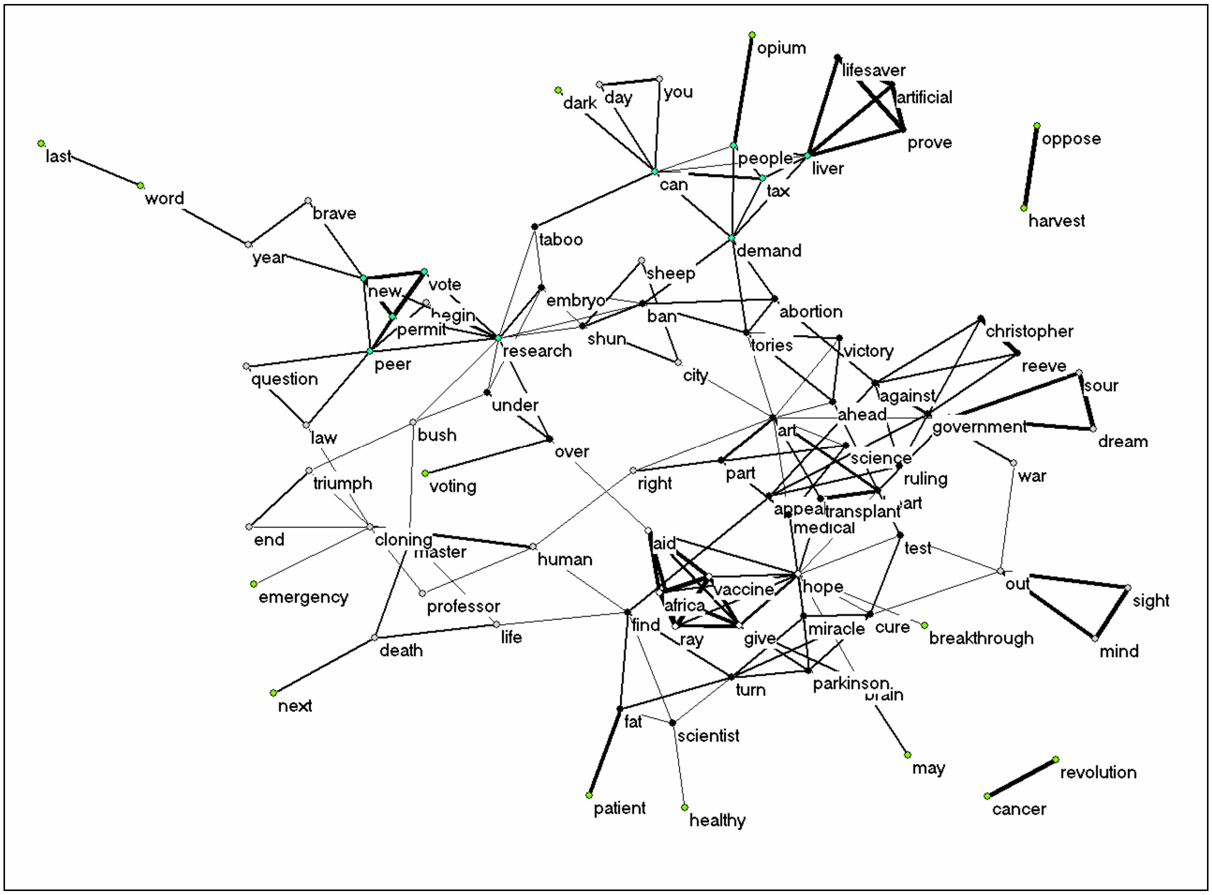

Figure 13. Map of 91 headline words in *The Guardian* 2001; cosine ≥ 0.2.

As in the case of the *New York Times*, these headline words are structured in terms of topics. The human dimension—for example, the hope for a cure of Parkinson's disease—is more prominent than the political issues. Normalization made the structure more informed, but did not affect the structure of the clusters. Six factors explain 22.2% of the variance (against 27.5% for the *New York Times*.) Thus, the newspapers headlines focus the attention in terms of topics, but not to the same degree as do the social sciences or the patent data. They allow for variation, but the topics are integrated into a single style by journalists and editors. The variation on the Internet or in comprehensive databases like the *Science Citation Index* and *MEDLINE* is therefore much larger.

7. Variation and Selection

The materials that we explored above are different mainly in the degree to which they exhibit variation or select from the variation using a specific codification. At the one extreme one would expect *AltaVista* as a representation of variation on the Internet. In this case the first six factors explained only 13.8% of the variance in the words. The landscape is very flat. All kinds of combinations occur, but some clusters constitute small islands in what is otherwise a sea of variation. Similarly, the *Science Citation Index* and the *MEDLINE* database provide a window on the semantic variation produced within the sciences. The structures are not more pronounced than in the case of *AltaVista*, but the islands are larger. Thus, the variation generating mechanisms in these two types of databases are different. The dedicated abstract and indexing services sustain a window on the variation generating mechanisms in the scientific discourses (that is, at the research front), while the Internet search engine provides a window on the variation without this additional dimension.

At the other extreme, the patent database organizes title words in a very precise manner so that the normalization and the factor analysis hardly added to the structure visible upon inspection of the observable co-occurrences. This was also the case in the *Social Science Citation Index*. In this latter database, the words are provided with specific meanings in the various contexts of scholarly debates. These meanings are differentiated along disciplinarily recognizable patterns. The newspapers take an in-between position. They also organize the variation by providing it with a specific interpretation, but the codification is less strict than in the two dedicated databases. In other words, the codification is differentiated less than in the dedicated databases. The language in newspapers is integrated differently from the codification in specialist jargons.

	Nr of documents retrieved	Nr of words included	Co-occurrence map	Map based on cosines	% variance explained by first 6 factors	% redundancy in the screeplot
USPTO	32	29	Very similar		61.0	19.7
SCI	2,634	156	Tight network	Network with islands	13.2	3.8
MEDLINE	1,597	71	Star shaped	Network with islands	17.9	4.6
Social SCI	53	41	Similar		57.7	22.2
AltaVista	1,060	119	Star shaped	Network	13.8	5.3
NYT	127	81	Star shaped	Rather similar	27.5	11.9
The Guardian	167	79	Star shaped	Rather similar	22.2	10.8

Table 1. A comparison of the databases in various dimensions

Because the percentage of variance explained by the first six factors is dependent on the number of variables (in this case, words) included in the analysis, we developed an indicator for the structuration of the word sets that is independent of the size of the word set. This indicator is added in the right-most column of Table 1. It indicates the redundancy in the distribution of the percentage of common variance explained by the subsequent factors (Sahal, 1979).[15]

The number of initial factors is by definition equal to the number of variables. The redundancy in the distribution of the percentages of variance explained can be expressed as a percentage of the maximum information content of this distribution.[16] This percentage redundancy provides us with a measure of the specificity of word usage at the network level.

The resulting column precisely confirms our more qualitative observations: the percentage of redundancy in the variation-producing databases remains under 5%, while it is approximately 20% for data from the *USPTO* database and the *Social Science Citation Index*. The two newspapers take in-between positions with 11 and 12% redundancy in the distribution of eigenvectors, respectively.

8. The mechanism of codification

Where does this lead us with respect to our initial question about the transfer of meaning by words functioning as metaphors that may carry the translation between contexts to a variable degree? The picture that has emerged hitherto is one of a source sending the information as variation to other databases which use topics and/or disciplinary delineations as focusing devices. Can we also make the translation between the different contexts visible? Let us now proceed by trying to specify this mechanism.

We focused above on each domain separately and in a comparative mode. However, one would need a single representation of two domains for studying the translation between them. For example, some texts contain *references* with title words that can be used as a representation of the knowledge base of the respective representations. The title words of these references can be mapped on the title words of the citing texts. The mutual information between these two domains (cited and citing texts) can then be disaggregated in terms of how much each word contributes to the translation.

For example, the 32 patents used above contain 584 references to 'non-patent literature' of which 409 provide us with titles of scientific documents that are articles in scientific journals or chapters in edited books.[17] These 409 cited titles contain 955 unique words. By using a threshold at a minimum of nine occurrences, this set was reduced to 88 words. These 88 words can be mapped against the 31 title words that we used above (Figure 4) for the mapping of the patent title words.[18] The two types of words can be organized into an asymmetrical matrix, and this can

[15] This distribution can be visualized in SPSS by asking for a scree plot in the factor analysis.

[16] The redundancy is defined as the difference between the maximum information content (H_{max}) of a distribution and its expected information content (H). The percentage redundancy is then equal to 100 * (H_{max} – H)/H_{max}. In this formula H is equal to $-\sum_i p_i \log(p_i)$ where $\sum_i p_i$ represents the distribution, and H_{max} is equal to the logarithm of the number of categories. This measure is independent of the number of variables (*n*) because $H_{max} = \log(n)$.

[17] See Leydesdorff (2004) for methodological details about the processing of the references in patent data.

[18] 29 title words were used in the previous analysis because the words 'stem' and 'cell' were deliberately excluded. In this case, however, these two words can be expected to contribute to the 'translation' of meaning from one context to another.

again be analyzed and visualized. The visualization (Figure 14) shows how the knowledge base of the scientific literature (cited) is represented in this set of patents (citing).

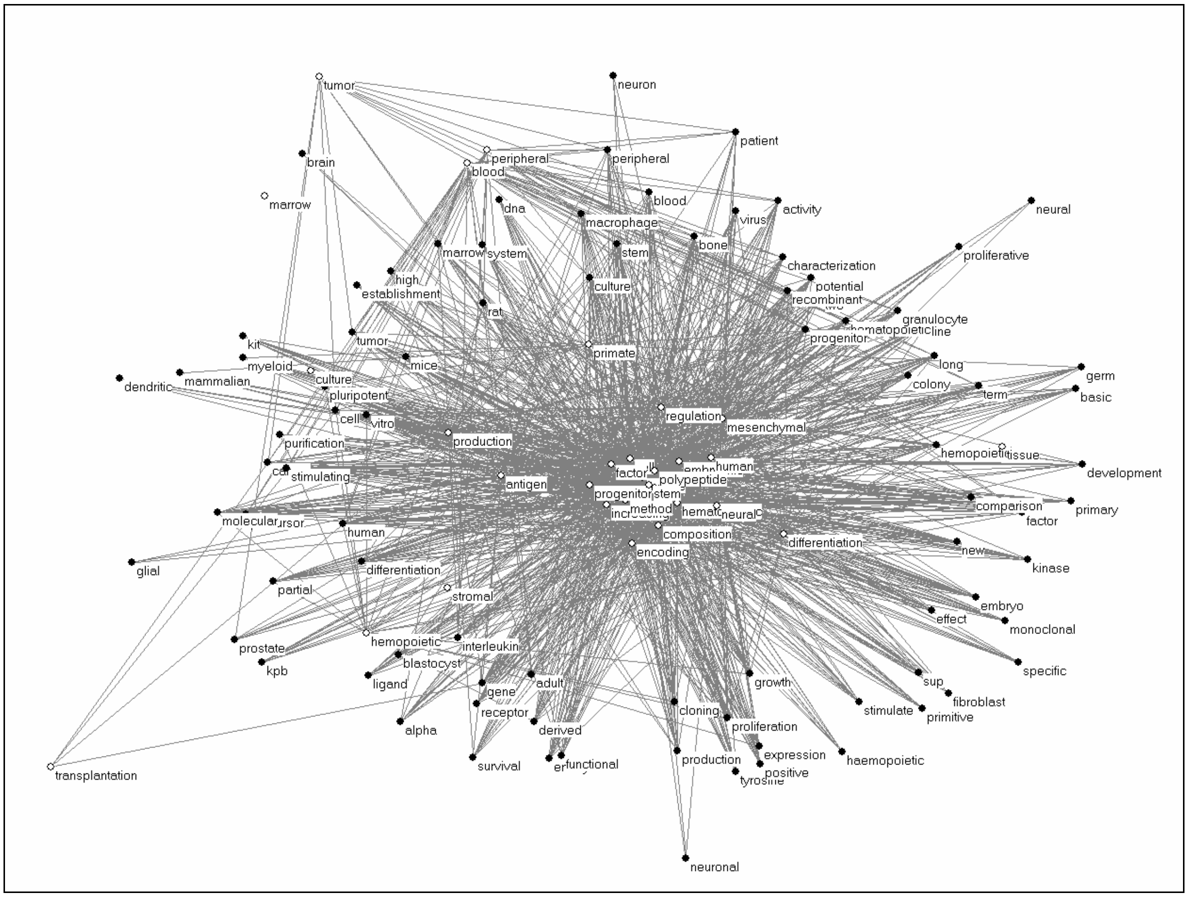

Figure 14. Map of the knowledge base of the patents (white vertices) in terms of the title words in their references (black vertices).

The resulting picture exhibits that the main group of patent words are centered and surrounded by the words from the scientific literature upon which they draw. At some places specific patent words (e.g., 'transplantation' and 'tumor') draw on a subset of this literature and thus structure the directionality in the mapping. In general, however, the patent words are surrounded by words from the relevant literature. The patents thus function as a focusing device within a knowledge base. The communication channels between these two literatures can be analyzed precisely by using words which occur in both sets.

The same technique can be applied to scientific journal literature insofar as this literature is increasingly available as full-text online, since the full texts include the respective lists of references. For the construction of the representation in Figure 15 we used the journal *Bone Marrow Transplantation* because this is the leading journal of this specialty.[19] Using the online

[19] Using the *Journal Citation Reports* the journal can be shown to be central to the specific cluster containing also all journals with 'stem cell' in their titles. The other major journal of this group is *Blood*, but this journal is less specifically focused on stem-cell research and its therapeutic applications.

edition of the journal, 181 papers could be retrieved from the volume of 2001. We parsed from these documents both the title words and the words used in the 4,740 scientific references in these papers. The 181 titles contained 697 unique words, of which 127 were used with an occurrence of three or more times (including ‘stem’ and ‘cell’). The 4,740 scientific references contained 4,016 unique words, of which 126 occurred 60 or more times.

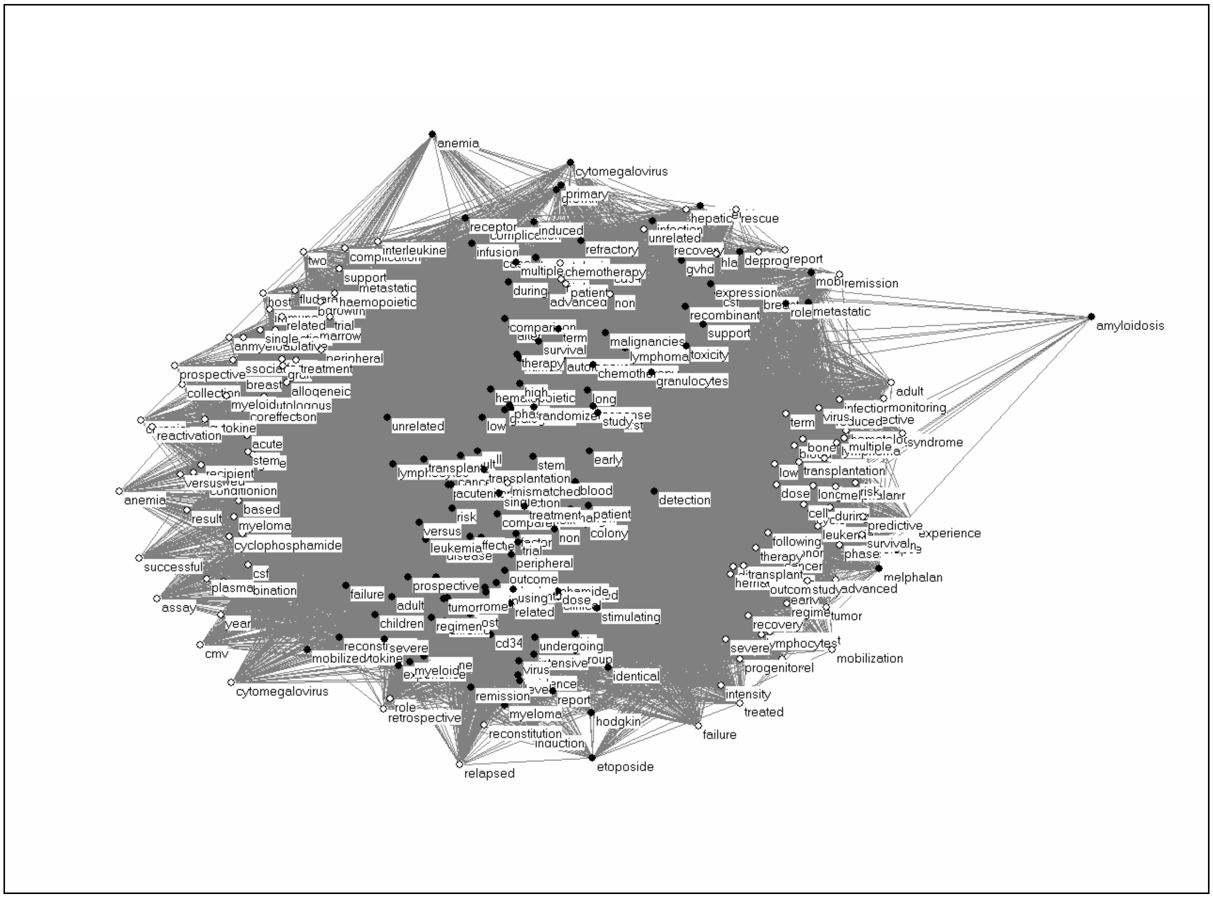

Figure 15. 127 title words versus 126 words occurring more often than 60 times in the 4016 scientific references of 181 articles in *Bone Marrow Transplantion* 2001.

The resulting picture exhibits a pattern *opposite* from the previous one: the title words of the references (black vertices) are now central to the network, and the title words of the citing articles provide the corona of the variation (white vertices). The references provide the common knowledge base for these articles on the basis of which the new variation is generated in different (that is in this case, mainly two) directions.

The use of the word combinations is again highly specific, and this specificity seems even higher than in the case of the USPTO data. The first factor of the words in the titles of the citing documents explains more than 70.0% of the variance of the bimodal matrix, while in the transposed case this increases to 86.6%. Table 2 summarizes these values and provides also the redundancy measure that was derived above for these four cases. The table shows that the specificity is by far the highest among the title words of the citing patents (41.5%). Patent citations to non-patent literature are carefully selected by the authors of patent applications and by the patent examiners.

	% variance explained by the first factor	% redundancy
USPTO Title Words	66.3	41.5
USPTO References	84.9	14.6
BMT Title Words	70.0	27.8
BMT References	86.6	14.9

Table 2. Skewness of the distribution of eigenvectors in the case of bimodal matrices of title words from citing and cited documents.

As noted, the channels that carry the translation between the cited and the citing dimensions can be decomposed in terms of the words present in both lists used for the analysis. For example, in the asymmetrical matrix between the 31 patent words (including ‘stem’ and ‘cell’) and 88 title words most frequent in the references of these patents, only 14 words are included both in the titles of the patents and in the titles of their scientific references. These words occur 624 times, that is, 6.2% of the total of co-occurrences of words (N = 10,114) in this matrix. The corresponding percentage for *Bone Marrow Transplantation* is 4.0% so that we may conclude that the communication between the citing and the cited dimension in terms of the title words involved is very selective. However, the words used on both sides of the interface are frequently used in both literatures and can thus be considered as fulfilling the function of communication channels or topical metaphors. The vast majority of the words, however, do not carry the translation. They are domain specific and thus contribute diaphorically to the organization of meaning.

9. Conclusions

The movement of information between scientists and the lay public is fraught with the potential for both positive impact and negative controversy. This study has attempted to consider the role of language in tracking the way that “stem cells” are represented in the patent literature, scientific articles, and newspapers through the quantification of word occurrences and the analysis of the structure in the network among them. The study was inspired by Luhmann’s (1984/1995) argument that social systems self-organize communications in terms of the meaning that is attributed to the communications. The meaning is no longer integrated in a central instance (e.g., the public), but increasingly differentiated into subsystems. Differentiation enables the communication to process more complexity. However, this raises the question of how communication is then possible at interfaces between the different domains of communication. What is the role of variation? Organized novelty production in the sciences, for example, cannot be understood without paying systematic attention to the uncertainty generating mechanisms at interfaces with relevant environments (Whitley, 1984; Leydesdorff, 2003; Fry, 2005).

In order to operationalize this question about how meaning is attributed to information and then also further communicated, we first have to be able to indicate meaning in the communication in terms of measurable units of analysis. In science studies and particularly in the sociology of translation, one has suggested that co-occurrences and co-absences of words can be used for mapping meaning in the dynamics of the sciences (Hesse, 1980; Callon *et al*., 1983, 1986; Law & Lodge, 1984). The measurement of meaning has thus progressed from using scales for measuring so-called semantic differentials (Osgood *et al*., 1957; Mitroff, 1974) to the information contained in the distribution of observable units in the data like words and co-occurrences of words. This operational definition of meaning, as a semantic field defined by the relations among words in a domain, makes it possible to study how communication is differently organized and codified in different contexts (Hesse, 1980; Leydesdorff, 1995, 1997).

Recent developments in visualization techniques enable us to study these different organizational formats both in terms of numerical information (e.g., factor-analytical results) and in terms of visualizations that allow for an appreciation. The main finding of this research was that the domains differ not only in terms of how the words are organized, and thus provided with meaning in relation to one another, but also in terms of the underlying processes which generate structures in the communication. Two types of processes were distinguished: those which primarily provide variations and those which have more the function of structuring the information. The various databases reflect these processes with different foci.

In our opinion, the most remarkable finding was that the *Social Science Citation Index* operates upon domains that are already structured as heavily as those of patent data. Both databases reflect the organization of word occurrences in terms of next-order categories like disciplinary structures and patent classifications. Internet search engines were expected to provide us with the widest form of variation, notably including occurrences within the common language. However, we found that the redundancy in the scientific databases that systematically provide variation (that is, the *Science Citation Index* and *MEDLINE*) was even somewhat lower than in the data generated from *AltaVista*. These scientific databases can perhaps be considered as specialized in representing the variation that is generated at the various research fronts.

Newspapers can be expected to provide their readers with 'story lines' that reduce uncertainty. Thus, the representations can be expected to operate on the selection side. The variation is integrated under common headlines. The *USPTO* and the *Social Science Citation Index*, however, were almost twice as selective as the newspapers under study. These databases can represent different disciplinary and technological frameworks, while the newspapers derive their identity from how they code the information into a single format. The variation in these next-order codification schemes provides a differentiated codification with more capacity than in the case of a single regime. In our opinion, this illustrates Luhmann's point about the function of differentiation in the processing of meaning (Leydesdorff, 2005). The meanings of communications are empirically contingent, variably codified, and to different extents. The methodology developed in this study enabled us to map the dynamics of these processes at the supra-individual level.

9. Further perspectives

The study of bi-modal matrices of title words in documents versus title words in the references of these documents has taught us that the references can be considered as highly specific channels of communication. The visualizations confirmed the hypothesis which was gradually developed on the basis of our finding that the patents carefully select from the scientific literature when citing it, while the scientific journal literature (in the natural and life sciences) can be expected to use the cited references to generate variation in new knowledge claims. At the research front, new meanings proliferate by making distinctions, while metaphors serve the integration of meaning across domains for individuals and institutions. The social system can be expected to develop its knowledge bases further in terms of such a variety of trade-offs between differentiation and integration.

The techniques of mapping debates and controversies about topics in several contexts systematically revealed the differences between the domains under study. Thus, we became informed about the multiple discourses used within and across the various domains. This systematic view of the data can be used as a basis for more qualitative and focused case studies on specific aspects of communication between the scientific and non-scientific discourses. One can nowadays process bodies of electronically available literature which are beyond the reach of more traditional forms of content analysis. The software needed for the visualization (Pajek) was already in the public domain (at <http://vlado.fmf.uni-lj.si/pub/networks/pajek>). We developed additional software that makes it possible to use document sets (titles and/or full texts) in a format that can be used directly as input for these mappings (at <http://www.leydesdorff.net/software/fulltext>). The source code is available from the first author for those readers who wish to develop these techniques further for academic purposes.

Acknowlegement

The authors are grateful to the suggestions of a number of anonymous referees.

References

Ahlgren, P., B. Jarneving, and R. Rousseau. 2003. Requirement for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient. *Journal of the American Society for Information Science and Technology* 54(6):550-560.

Bar-Hillel, Y. 1955. An Examination of Information Theory. *Phil. Sci.,* 22:86-105.

Bar-Ilan, J. 2001. Data collection methods on the Web for informetric purposes--A review and analysis. *Scientometrics,* 50(1): 7-32.

Bhattacharya, S., H. Kretschmer, & M. Meyer. (2003). Characterizing Intellectual Spaces between Science and Technology. *Scientometrics*, 58(2) : 369-390.

Bernstein, B. 1971. *Class, Codes and Control, Vol. 1: Theoretical Studies in the Sociology of Language*. London: Routledge & Kegan Paul.

Biagioli, M., & P. Galison, eds. 2003. *Scientific Authorship: Credit and Intellectual Property in Science.* New York: Routledge.

Burt, R. S. 1982. *Toward a Structural Theory of Action*. New York: Academic Press.

Callon, M., J.-P. Courtial, W. A. Turner, and S. Bauin. 1983. From Translations to Problematic Networks: An Introduction to Co-word Analysis. *Social Science Information* 22:191-235.

Callon, M., J. Law, and A. Rip, eds. 1986. *Mapping the Dynamics of Science and Technology.* London: Macmillan.

Coser, R. L. 1975. The Complexity of Roles as a Seedbed of Individual Autonomy. In *The Idea of Social Structure. Papers in Honor of Robert K. Merton,* edited by L. A. Coser, 237-264). New York/Chicago: Harcourt Brace Jovanovich.

Fry, J. 2005. Scholarly Research and Information Practices: A Domain Analytic Approach. *Information Processing and Management* 42:(forthcoming).

Gieryn, T. F. 1999. *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University of Chicago Press.

Gilbert, G. N., and M. J. Mulkay. 1984. *Opening Pandora's Box. A Sociological Analysis of Scientists' Discourse*. Cambridge: Cambridge University Press.

Granstrand, O. 1999. *The Economics and Management of Intellectual Property: Towards Intellectual Capitalism*. Cheltenham, UK: Edward Elgar.

Grupp, H., & U. Schmoch. 1999. Patent Statistics in the Age of Globalisation: New Legal Procedures, New Analytical Methods, New Economic Interpretation. *Research Policy* 28: 377-396.

Guston, D. 2000. *Between Politics and Science.* Cambridge, UK, etc.: Cambridge University Press.

Habermas, J. 1981. *Theorie Des Kommunikativen Handelns.* Frankfurt a.M.: Suhrkamp.

Hellsten, I. 2002. *The Politics of Metaphor* (Vol. 876). Tampere: University of Tampere; at <http://acta.uta.fi/pdf/951-44-5380-8.pdf>.

———. 2003. Focus on Metaphors: The case of 'Frankenfood' on the web. *Journal of Computer-Mediated Communication* 8(4) July; at <http://www.ascusc.org/jcmc/vol8/issue4/hellsten.html>

Hellsten, I., & L. Leydesdorff. 2004. *Measuring the Meanings of Co-Words In Contexts: Automated Analysis of 'Monarch Butterflies', 'Frankenfoods', and 'Stem Cells'.* Paper presented at the Conference of Research Council 33 of the International Sociological Association, 17-21 August 2004, Amsterdam.

Hesse, M. 1980. *Revolutions and Reconstructions in the Philosophy of Science*. London: Harvester Press.

———. 1988. The Cognitive Claims of Metaphors. *Journal of Speculative Philosophy,* 2(1):1-16.

Jaffe, A. B., and M. Trajtenberg. 2002. *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA/London: MIT Press.

Jasanoff, S. 1990. *The Fifth Branch: Science Advisers as Policymakers*. Cambridge, MA: Harvard University Press.

Kamada, T., and S. Kawai. 1989. An algorithm for drawing general undirected graphs. *Information Processing Letters* 31(1): 7-15.

Kuhn, T. S. 1984. Scientific Development and Lexical Change. *The Thalheimer Lectures* Johns Hopkins University.

Law, J., and P. Lodge. 1984. *Science for Social Scientists*. London, etc.: Macmillan.

Lewis, R. 1997. Embryonic stem cells debut amid little media attention. *The Scientist,* 11(19): 1-2.

Leydesdorff, L. 1989. Words and Co-Words as Indicators of Intellectual Organization. *Research Policy* 18:209-223.

———. 1993. Why the Statement “Plasma-Membrane Transport Is Rate-Limiting for Its Metabolism in Rat-Liver Parenchymal Cells” Cannot Meet the Public. *Public Understanding of Science* 2:351- 364.

———. 1995. *The Challenge of Scientometrics: The Development, Measurement, and Self-Organization of Scientific Communications*. Leiden: DSWO Press, Leiden University; at < http://www.upublish.com/books/leydesdorff-sci.htm >.

———. 1997. Why Words and Co-Words Cannot Map the Development of the Sciences. *Journal of the American Society for Information Science* 48 (5):418-427.

———. 2001a. *A Sociological Theory of Communication: The Self- Organization of the Knowledge-Based Society*. Parkland, FL: Universal Publishers; at <http://www.upublish.com/books/leydesdorff.htm>.

———. 2001b. Indicators of Innovation in a Knowledge-based Economy. *Cybermetrics* 5(1), Paper 2, at <http://www.cindoc.csic.es/cybermetrics/articles/v5i1p2.html>.

———. 2003. The Construction and Globalization of the Knowledge Base in Inter-Human Communication Systems. *Canadian Journal of Communication* 28(3):267-289.

———. 2004. The University-Industry Knowledge Relationship: Analyzing Patents and the Science Base of Technologies. *Journal of the American Society of Information Science & Technology* 55(11):991-1001.

———. 2005. Meaning, Anticipation, and Codification in Functionally Differentiated Social Systems. In *Luhmann simulated – Computer Simulations to the Theory of Social Systems*, edited by Th. Kron, U. Schimank, & L. Winter. Münster, etc: Lit Verlag (forthcoming).

Luhmann, N. 1984. *Soziale Systeme. Grundriß einer allgemeinen Theorie*. Frankfurt a. M.: Suhrkamp. [Stanford: Stanford University Press, 1995].

———. 1990. The Cognitive Program of Constructivism and a Reality that Remains Unknown. In *Selforganization. Portrait of a Scientific Revolution,* edited by W. Krohn, G. Küppers and H. Nowotny, 64-85. Dordrecht: Reidel.

———. 1996. *Die Realität der Massenmedien*. Opladen: Westdeutscher Verlag [Stanford: Stanford University Press, 2000].

Maasen, S., and P. Weingart. 1995. Metaphors—Messengers of Meaning. *Science Communication* 17(1):9-31.

McInerney, C., N. Bird, and M. Nucci. 2004. The Flow of Scientific Knowledge from Lab to the Lay Public. *Science Communication 26*(1): 44-47.

McLuhan, M., Q. Fiore, and J. Agel. 1969. *The Medium Is the Massage*. Hormondsworth, etc.: Penguin.

Mitroff, I. I. 1974. *The Subjective Side of Science*. Amsterdam: Elsevier.

Mulkay, M., J. Potter, and S. Yearley. 1983. Why an Analysis of Scientific Discourse Is Needed. In *Science Observed: Perspectives on the Social Study of Science*, edited by K. D. Knorr and M. J. Mulkay. 171-204. London: Sage.

Narin, F., and D. Olivastro. 1988. Technology Indicators Based on Patents and Patent Citations. In *Handbook of Quantitative Studies of Science and Technology,* edited by A. F. J. v. Raan, 465-507. Amsterdam: Elsevier.

Nisbet, M. C., D. Brossard, and A. Kroepsch. 2003. Framing science: The Stem Cell Controversy in an Age of Press/Politics. *Press/Politics 8*(2): 36-70.

Nowotny, H., P. Scott, and M. Gibbons. 2001. *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge: Polity.

Osgood, C. E., G. Suci, and P. Tannenbaum. 1957. *The Measurement of Meaning*. Urbana: University of Illinois Press.

Rousseau, R. 1999. Daily time series of common single word searches in AltaVista and NorthernLight,. *Cybermetrics 2/3*, Paper 2 at <http://www.cindoc.csic.es/cybermetrics/articles/v2i1p2.html>.

Sahal, D. 1979. A Unified Theory of Self-Organization. *Journal of Cybernetics* 9:127-142.

Salton, G., and M. J. McGill. 1983. *Introduction to Modern Information Retrieval*. Auckland, etc.: McGraw-Hill.

Small, H. 1999. Visualizing Science by Citation Mapping. *Journal of the American Society for Information Science* 50(9):799-813.

Small, H., and E. Greenlee. 1986. Collagen Research in the 1970s. *Scientometrics* 19(1-2):95-117.

Thelwall, M. 2001. The Responsiveness of Search Engine Indexes. *Cybermetrics*, 5(1), at <http://www.cindoc.csic.es/cybermetrics/articles/v5i1p1.html>.

Weelwright, P. 1962. *Metaphor and Reality*. Bloomington: Indiana University Press.

Wertz, D. C. 2002. Embryo and stem cell research in the USA: a political history. *Trends in Molecular Medicine* 8(3):143-148.

Whitley, R. D. 1984. *The Intellectual and Social Organization of the Sciences*. Oxford: Oxford University Press.

Wouters, P., Hellsten, I. & Leydesdorff, L. 2004. Internet time and the reliability of search engines. *First Monday 9*(10) at http://www.firstmonday.org/issues/issue9_10/wouters/index.html