Vol 5 Issue 2 Aug 2015 ISSN No :2231-5063

International Multidisciplinary Research Journal

Golden Research Thoughts

Welcome to GRT

RNI MAHMUL/2011/38595 **ISSN No.2231-5063**

Golden Research Thoughts Journal is a multidisciplinary research journal, published monthly in English, Hindi & Marathi Language. All research papers submitted to the journal will be double - blind peer reviewed referred by members of the editorial board.Readers will include investigator in universities, research institutes government and industry with research interest in the general subjects.

International Advisory Board

Editorial Board

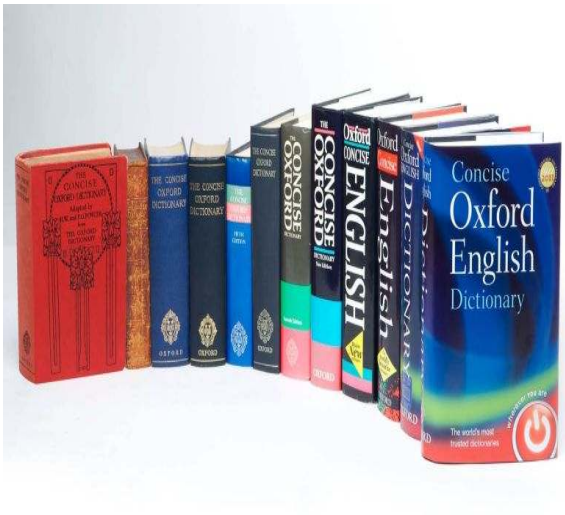
Address:-Ashok Yakkaldevi 258/34, Raviwar Peth, Solapur - 413 005 Maharashtra, India
Cell : 9595 359 435, Ph No: 02172372010 Email: ayisrj@yahoo.in Website: www.aygrt.isrj.org

Golden Research Thoughts
ISSN 2231-5063

Impact Factor : 3.4052(UIF)
Volume - 5 | Issue - 2 | Aug - 2015

THE CONTEXT – SENSITIVE REPRESENTATION OF MEASURING THE MEANING OF WORDS AND WRITING ASSISTANCE TO THE NEXT GENERATION: A STUDY OF ENGLISH LANGUAGE DICTIONARY

Sudhansu Dash



ABSTRACT

The term 'dictionary' loosely denotes a wide range of reference sources useful to students across all disciplines. It is often felt to apply mainly to works giving linguistic information about words, such as their spelling, pronunciation, grammatical class, meanings, phrasal and collocational combinations, related words, and varietal restrictions. By contrast 'encyclopaedias' contain detailed factual, cultural and other non-linguistic information. However, in practice, many works named dictionaries contain the latter information in considerable quantities - general encyclopaedic dictionaries, and specialist subject area dictionaries such as dictionaries of biography, architecture, civilization, literature, politics and indeed of languages and linguistics.

KEYWORDS :*biography, architecture, civilization, literature, politics .*

INTRODUCTION :

Among more language focused dictionaries we may distinguish those that are bilingual (or indeed multilingual/polyglot) from those that are monolingual. The organization of the words treated is usually by alphabetical order, though it may also be 'thematic' - in meaning-related groups - in which case the work may be called not a dictionary, but a 'thesaurus' or 'word finder' Furthermore, many dictionaries specialize either in specific types of word (e.g. idioms, place names, abbreviations, words commonly confused) or in words of specific registers/varieties/historical periods (e.g. slang, medical English, architectural terms, Old English, American English, Shakespeare's English) or indeed in specific aspects of lexical information e.g. pronouncing, combinatory, synonym, valency, usage, or etymological dictionaries.

Reference works of all these types no longer appear solely in book form, but also electronically as hand-held devices, on CD and the internet Furthermore, they come in forms designed for various purposes and users: in particular they may be more scholarly, designed mainly as academic records or more practically designed as aids to help native speakers, translators or learners/students when they have word-related problems. In turn the latter may be designed for specific ages or levels of language

ability. While the issues we take up below apply in principle to all of the above, they have mainly been explored with respect to dictionaries for language learners.

DICTIONARY DESIGN

Years ago, dictionary writers relied heavily on their own expert intuitions, prompted by existing works. Some empirical element was provided by consulting collections of cards of citations, on which interesting examples of bits of text containing words used in new or interesting ways had been recorded. Often these were gradually built up not only by lexicographers but also by members of the general public, employed to 'read and mark' particular types of written material (e.g. daily newspapers). The main focus of the dictionary maker was on factual correctness, and on recording every strange detail of a word's behavior. However, recently there has been something of a revolution in dictionary making, exhibiting three notable trends.

Use of insights from Linguistics

The increasing involvement of expertise in linguistics appears in such matters as the distinguishing and coding of complementation patterns of verbs, the semantic analysis of words into components as an aid to creating suitable definitions or thematic organizations, and the delimitation of different categories of phrases and collocations displayed using different fonts.

Use of electronic data

An increasing reliance on electronic corpora, pioneered by the COBUILD project is now widely adopted. This has moved the collection of information away from the narrow sources described above and potentially makes available a colossal amount of authentic linguistic data to draw on when composing entries. Standard corpora of general current English such as the British National Corpus contain 100 million running words of which, significantly, 10% is from tapes of spoken language, much of it spontaneous conversation. Such resources provide two important types of information

One can instantly retrieve all, or a sample of, the occurrences of any word or phrase of interest together with as much of the surrounding text as one wants (i.e. a concordance): the standard retrieval is 'keyword in context' (KWIC) where corpus lines containing a word of interest are displayed on screen one above the other, with the word of interest centered and highlighted. This is an invaluable resource not only to prompt the lexicographer to refine all types of lexical information, but also as a source of authentic examples to be included, and large numbers of these can be accommodated in the new electronic versions of dictionaries.

They provide accurate frequency information both on words and (with some extra analytic effort from the lexicographer, or the involvement of computational linguists) on particular senses and phrasal combinations of words, or the occurrence of words with different complementation patterns (e.g. like + v-ing, versus like+ to v). This can be obtained for the language as whole or specific varieties such as conversation, academic writing etc. and enables dictionaries to include accurate frequency information for the user as well as inform the selection and internal organization of entries.

User-friendliness

The increasing trend concerns awareness of the user. With the rise of a learner-centered view of learning, researchers have drawn lexicographers' attention to the need for most dictionaries not just to give unassailably correct information, but also to present it in ways that the targeted user can easily exploit successfully. Though some such issues go back a long time (e.g. attention to the need for

learners' dictionaries to use a limited defining vocabulary so as to avoid the problem of a definition being harder than the word defined), there is much more attention these days to general matters of user-friendliness, such as: cross-referencing in order to help users find the information they need from a variety of look-up starting points; including (in electronic versions) sound and video clips; minimal use of codes for grammatical features, and of a type the user is likely to be familiar with; making sure examples do not contain unnecessary complexities; using a variety of styles of word definition; focusing on the standard authentic uses of words more than the rare and exceptional ones; using frequency information to select and order the information in entries so that the most frequent will be encountered first in the entry (which may override the tradition of entering information on phrasal uses of words always after the uses of the word in isolation).

Measuring the Meaning of Words in Contexts:

Co-words have been considered as carriers of meaning across different domains in studies of science, technology, and society. Words and co-words, however, obtain meaning in sentences, and sentences obtain meaning in their contexts of use. At the science/society interface, words can be expected to have different meanings: the codes of communication that provide meaning to words differ on the varying sides of the interface. Furthermore, meanings and interfaces may change over time. Given this structuring of meaning across interfaces and over time, we distinguish between metaphors and diaphors as reflexive mechanisms that facilitate the translation between contexts. Our empirical focus is on three recent scientific controversies: Monarch butterflies, Frankenfoods, and stem-cell therapies. This study explores new avenues that relate the study of co-word analysis in context with the sociological quest for the analysis and processing of meaning.

Skills and use

Dictionary use has only relatively recently becomes a topic of research interest Studies divide into (a) questionnaire surveys concerning what dictionaries people (say they) use, how often, what they look up in them, and the like, and (b) research on the detailed skills or strategies that users possess, or need to possess, when actually consulting dictionaries for various specific purposes, often using diary, interview or think aloud research methods. Though we can only tentatively generalize across all kinds of user and situation, current research suggests that dictionaries are used about equally in the process of writing, and sometimes when just studying/learning. Spelling and meaning is the information most commonly looked up, with much valuable information in entries (e.g. about grammar and collocation) being underexploited. Good students often draw on more than one dictionary, and they progress from reliance on bilingual to monolingual target language dictionaries. In the reading process, dictionary use competes with various kinds of guessing, or just ignoring unknown words that come up. There is strong evidence that expert readers make good choices about when to use each of these, do not use the dictionary exclusively, and often do so after making attempts at guessing. Various necessary sub skills have been identified in dictionary lookup, such as rapid alphabetical order search, readiness to check in more than one place for an apparently missing word, ability to scan and select from a polysemous entry. In writing, the dictionary may be called upon for a wide range of types of information besides word meaning. Often a writer retrieves a word for what they want to express, but needs to check some aspect other than its meaning (e.g. irregular verb tense form, or what a typical object might be), or choose between two words they have retrieved. Dictionary use either in reading or writing may lead on to learning, and may additionally be used along with more decontextualised learning strategies as a learner memorizing wordlists in a foreign language may 'resource' from the

dictionary to check information on a word, or to browse the entry for further meanings and information to master.

Training and assessment

Training in the use of dictionaries, though ostensibly an important aspect of study skills training, is not universally accepted or widely practiced, and indeed is a weakness in the training of teachers themselves .Language teachers have often regarded dictionary use negatively, taking the view that it encourages laziness (the learner should make the effort to guess unknown words) or that it distracts a class's attention from the teacher or, where bilingual dictionaries are involved, that it leads to unwanted 'thinking in the first language'. Consequently students' dictionary skills are often poor.

Review of Literature

A large number of texts can be retrieved from the Internet for research purposes through the use of search engines; citation index databases, on-line archives of newspapers, scientific journals, popular scientific magazines, and on-line discussion groups; the websites and databases of various governmental, non-governmental, and commercial organizations can also be mined. This overload of textual materials poses new methodological challenges for the disciplines in the humanities and social sciences that are interested in text analysis. How can one automate the analysis of large amounts of texts that can no longer be analyzed qualitatively or coded manually, and still obtain conceptually meaningful and valid results?

Several research traditions, such as computer-aided content analysis, corpus-based linguistics, and the so-called 'sociology of translation' (Callon et al., 1986; Stegman & Grohmann, 2003) have developed tools for the automated analysis of texts. Despite the different disciplinary backgrounds and research agendas of these traditions, they have all faced similar problems with the ambiguity of language. Words and the relations among words ('co-words') mean different things in other contexts, and the meaning of words can be expected to change, particularly in science where novelty production is part of the mission of the enterprise (Whitley, 1984; Fry, 2006). Without further reflection, words and co-words cannot be used for mapping the dynamics of science and technology (Leydesdorff, 1992; 1997).In other words, one needs to specify a next-order mechanism of meaning exchanges to study both the changing distributions of words and the variance in their meanings and relations. For example, Luhmann (1984, 1986) argued that social systems communicate by processing meaning on top of information exchanges. From this perspective, meaning-processing is considered as a property of the systems of coordination in society. Meaning is generated by positioning the communication within networks of relations. Thus, meanings can be expected to vary across domains of use (e.g. science, journalism, economics). For example, one can generate value in economic transactions, but scientific theories are improved through discursive arguments. While Luhmann focused on the differentiation of meaning-processing and was not so much interested in the relation between meaning-processing and information-processing, this interface is precisely the challenge for the information scientist. We are interested in whether meanings can be traced and measured in the communications that occur between the different domains of use (such as the sciences, the economy, and the mass media) of society, and whether such mappings can be automated. In computer-aided content analysis (e.g. Klein, 2004) the main focus has been on processing large bodies of textual data and on automatically coding specific aspects of the texts. Searching for particular words in documents, creating word frequency lists, and listings of word concordances have been automated, but within this tradition the coding schemes have to be developed by the analysts. As Krippendorff (1980/2002) notes, there remains a

need to specify the context in which the texts become meaningful. Similarly, the main aim of research in corpus-based linguistics (e.g., Kennedy, 1998) has been to automate corpus analysis by tagging words within their grammatical contexts and clustering various tokens of the same word (e.g. 'word' and 'Words') as belonging to the same type. Yet, the problem of the semantic ambiguity of words has remained.In science and technology studies, co-occurrences of words ('co-words') have been considered as the carriers of meaning across different domains (Callon et al., 1983). In the so-called 'sociology of translation' (Callon et al., 1986; Callon et al., 1991), co-words have been used to map the dynamics of science and technology in terms of translations. The main focus in quantitative studies of translations has been on the network of co-occurring key words as indicators of activity in the document sets (Callon et al., 1991; Ruiz-Baños et al., 1999; Stegmann & Grohmann, 2003; Bailón-Moreno et al., 2005). In the network of co-words, however, "the robustness of structured relations does not depend on qualities inherent to those relations but on the network of associations that form its context" (Teil & Latour, 1995).Stegmann & Grohmann (2003) emphasized that co-words are particularly suited for the study of 'weak links' (Granovetter, 1973): the co-words relate otherwise unconnected literatures. These authors proposed to call this activity 'Swanson Linking' because in a series of articles Swanson (e.g., 1990, 1999) used this linking for discovering new relations like adverse drugs reactions (Rikken, 1998; cf. Rikkenet al., 1995). Our approach differs from these studies in that our focus is not on the relations and co-occurrences of words, but on the positions of words in different semantic fields. These positions can be considered as the unintended results of a set of relations in a network among agents or documents (Burt, 1982, 1983).In other words, we are not only interested in dyadic co-occurrences, but also in single occurrences and triadic (etc.) co-occurrences. Accordingly, we will not use the co-occurrence matrix but the underlying asymmetrical matrix of documents versus words, and subsequently compute the distance among the word vectors using the vector-space model, that is, using the cosine as a similarity measure. The co-occurrence matrix—which contains less information—can be obtained by multiplying the asymmetrical matrix with its transposed (Leydesdorff, 1989; Leydesdorff & Vaughan, forthcoming).

Our specific focus is on science communication because at the interface between science and other domains of society, words can be expected to have different meanings. These domains use different codes for the communication, and also the degree of codification may differ across the domain of use. For example, in daily life, a 'shortage of energy' means something very different from the concept of 'energy' as a conserved quantity in physics. The degree of codification of the words is higher in scientific articles than in the mass media. Furthermore, in the sciences, meanings can be expected to change with the development of new knowledge.

As case studies, we use three scientific controversies that have flourished recently in public debates: first, Monarch butterflies; second, Frankenfoods; and third, stem cells. However, before turning to these case studies, let us first discuss in more detail the problem of automating the mapping of the meanings of the words and the question of what could be considered as providing the contexts for such mapping.

Purpose of the study

There is increasing research into the dictionary strategies of the user: clearly there is a limit to what the dictionary can do to help the user and good dictionary skills need to be trained, though such training has often been neglected.

METHODOLOGY

The datasets are specified below in the three case studies separately, but the researcher utilised a common methodology in all three cases in order to reduce the complexity in the comparison. After the first case study that—as noted—focuses on the debate about the genetic modified corn pollens and the Monarch butterfly, the researcher scales up in a second step to sets of documents that can no longer be read and coded manually. To help the purpose the researcher draws upon two previous case studies in which techniques have been developed to trace mechanisms for reflection among textual domains. In one study, traced the metaphor of "Frankenfoods" model on the web over time and the other, Leydesdorff & Hellsten model used the diaphor "stem cells" to map words and co-words in contexts across different domains like newspapers, the Internet, and scientific databases. The techniques are based on commonly available software programs using the Internet module available in Visual Basic. In all case studies, the files were first parsed so that each document represented a separate text file. These documents were then broken down into sentences and words. Word frequency lists were generated. The researcher selected only the body text for analysis—in some cases the full text, and in the case of large sets the titles—thus excluding additional information included on the web pages. The maps were optimized for visualization using pragmatic cut-off levels of word frequencies in order to keep them readable. The researcher used approximately one hundred words as the maximum. Though technically, it is possible to include many more words in the analysis, but then the reading of the maps becomes problematic.

DELIMITATION OF THE STUDY

Both diaphors and metaphors can be studied diachronically and/or synchronically. In this study, the researcher limits the analysis to a diachronic discussion of the metaphor and a synchronic comparison in the case where the researcher expects a diaphor. Thus, the focus is on the two extreme poles of a continuum of potentially different mechanisms of codification. A very pronounced metaphor ("Frankenfoods") is studied in a largely un-codified set of documents, and a common word ("stem cell") in a set of codified texts. However, the researcher first validate our methodology by using a qualitative study of five documents central to the controversy about the potentially harmful effects of genetically modified corn pollen on Monarch butterflies. This case allows the researcher to build upon an argument that in the translation of science to various publics, the frames of reference are different in the various domains and their related discourses. Can the differences in meaning indicated by these authors be made automatically visible by using our methods?

Mapping translations between contexts: metaphors and diaphors

Information is codified when provided with meaning. Some meanings, more than others, gain resonance between the different domains in society. In the analysis of how meaning is given to the uncertainty contained in a distribution of words, one can distinguish between a diachronic problem and a synchronic problem. The synchronic problem is further complicated when different meanings—which can each be codified in different domains—are exchanged as in social systems. The synchronic and the diachronic mechanisms may further interact in a non-linear mode; meanings can then be stabilized locally and sometimes further be meta-stabilized and globalised, as in scientific communication.

Historically, the measurement of meaning has had two relatively independent roots. On the one hand, researchers have attempted to measure meaning from a psychological perspective using scales . On the other hand, information science research has focused on how the measurement of

meaning can be operationalised using words and their co-occurrences. The information-theoretical tradition is dominated by a semantic orientation on meaning as a structural property of the communication networks, while the psychological measurement can be considered as based primarily on a pragmatic theory of meaning. The researcher focuses in this study is on measuring meaning within the semantic tradition, that is, as a property of the network of words.

In order to specify the context for measuring meanings, the researcher focused on specific kinds of words, notably, words that can be considered as flagships for the debates. Specific terms such as "Frankenfoods" and "stem cells" are used in scientific and popular-scientific domains as well as in journalism, and may therefore provide common ground for the different discourses while still functioning differently in each of these domains. Such words as metaphors can be considered as 'messengers of meaning.' Metaphors would generate the dynamics of knowledge.

The expectation is that a metaphor can be considered as one reflexive mechanism in the networks of words among others. A metaphor can act as a 'messenger of meaning' or a 'translator spokesman' in a symbolic manner because its occurrence is punctuated. Translation, however, can also be sub-symbolic, that is, a result of the interactions among different densities in the network. Translation in science communication may thus function both symbolically (as metaphors) and sub-symbolically (as diaphors). The researcher hypothesizes that metaphors and diaphors can be considered as tools of intermediation that channel meanings across different arenas in the communication of science. This is possible because they both contribute to carrying a set of relations from one domain to another.

The concept of 'diaphor' makes an analytical distinction between words that carry meaning (i.e., metaphors), and words that contribute to the boundary construction between domains of communication in discourses (Weelwright, 1962). Whereas metaphors such as "Frankenfoods" can be considered as punctuated tools of intermediation that channel meanings among otherwise different semantic fields, common words such as "stem cells" obtain meaning from their positions in the field of relating words. A metaphor brings domains together in a symbolic mode, while common words are expected to function sub-symbolically; their contribution to the translation of meaning is the result of interactions among the various clusters and hubs in the networks of words on the different sides of an interface. In the sub-symbolic case, the tensions found in the meaning of these terms are not necessarily resolved.

RESULTS

Both diaphors and metaphors can be studied diachronically and/or synchronically. In this study,the researcher limits the analysis to a diachronic discussion of the metaphor and a synchronic comparison in the case where the researcher expects a diaphor. Thus, the focus is on the two extreme poles of a continuum of potentially different mechanisms of codification. A very pronounced metaphor ("Frankenfoods") is studied in a largely un-codified set of documents, and a common word ("stem cell") in a set of codified texts. However, the researcher first validate our methodology by using a qualitative study of five documents central to the controversy about the potentially harmful effects of genetically modified corn pollen on Monarch butterflies. This case allows the researcher to build upon an argument that in the translation of science to various publics, the frames of reference are different in the various domains and their related discourses. Can the differences in meaning indicated by these authors be made automatically visible by using our methods?

The different rhetoric used in these five documents to illustrate how scientific information is carried across media boundaries. The rhetorical changes altered the case and most likely served as a

catalyst for the media frenzy that accompanies the discussion. In order to test the methodology, the researcher first shows that by measuring the meanings of the (co-)words in these five documents, he able to visualize the rhetorical changes in different frames indicated in the figures below.

From our methodological perspective, the paragraphs in the five documents provide us with the cases to which the words are attributed as variables. Only eight (non-stop word) words are used in all five documents ('pollen,' 'corn,' 'monarch,' 'field,' 'butterfly,' 'feed,' 'grew,' and 'laboratory') and only two of these words, namely 'pollen' and 'monarch,' occur more than twice in each of the documents. The researcher focuses on these two words in order to show the change of the positions. In order to sort out how these words are positioned in the different documents, semantic maps are drawn using all the words that occur at least twice in a given document.

As these are single document studies, the cosine threshold for inclusion in the graph is set at the level of larger than or equal to 0.5 . The cosine values are affected by the density of the relations: the tighter the network, the higher the threshold has to be set in order to produce a map that exhibits the semantic organization. Unlike document sets, single documents provide 'restricted discourses' that one can expect to be well organized in word usage and tightly connected, while one can expect that 'elaborate discourses' among documents are more loosely organized .For this reason, the researcher shall use a threshold of cosine ≥ 0.1 in the case of large document sets.

In the semantic map that results (Figure 1), the two words that were our focus, namely 'pollen' and 'Monarch,' are part of different word clusters, thus illustrating how they embody different parts of the argument. In order to draw attention to the clusters that the researcher wished to focus on, we illustrate them with grey shades. The methodology of the research is visible as a third grouping. As expected in the case of scientific literature, the different parts of the argument are clearly separated from one another in terms of the cause, the effect of the problem, and the work process that validates the inference.

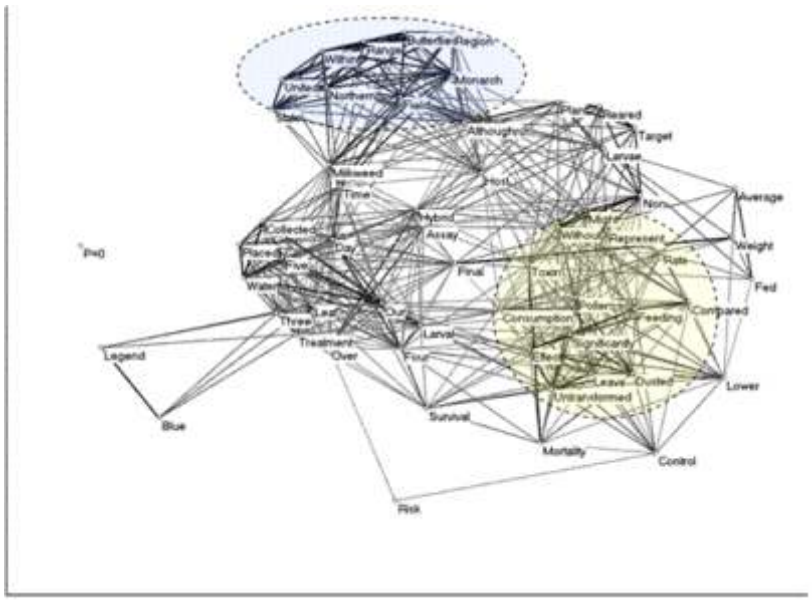


Figure 1: The cosine map of 59 words used more than once in the Scientific Correspondence published in Nature, 399: 214 on May 20, 1999 (cosine ≥ 0.5) .

Unlike the practice in corpus-based linguistics, the researcher did not group the tokens 'larvae' and 'larval' as a single type; in the figure they are grouped differently. In a six-factor solution of the matrix (which explains 94.2% of the variance), for example, 'larval' has a factor loading of 0.855 on factor two—mainly representing methodological words—while the word 'larvae' loads on the fifth factor with –0.758. (There is not a lot of inter factorial complexity in the orthogonally rotated solution.) The 'larvae' are among the subjects of study, while the word 'larval' belongs to the methods section of the argument. These distinctions are very sensitive in scientific literature .If these words would be grouped together in a coding scheme ex ante, the semantic map would have been distorted.

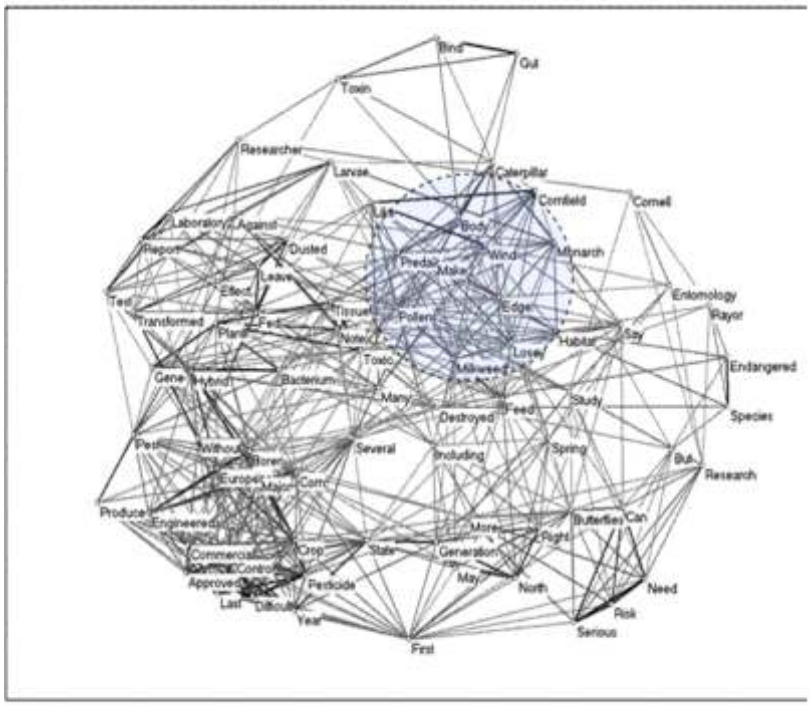


Figure 2: The cosine map of 77 words used more than once in the Cornell University press release released on May 19, 1999 (cosine ≥ 0.5).

The next picture (Figure 2) provides a similar representation of 795 words, of which 296 were unique. Seventy-seven words occurred more than once and were therefore included in the analysis. Unlike the earlier figure the main common words, 'pollen' and 'Monarch,' are here parts of the same word cluster. The argumentative structure of the scientific contribution is merged in this reflection with another purpose, notably to draw attention to the main findings of the researchers, and the possible implications of the findings are emphasized.

Furthermore, this map shows that the press release raised a new topic that relates to the European corn borer—against which the corn was genetically modified. Whereas Nature talked about 'larvae', the press release uses both the terms 'caterpillar' and 'larvae.' 'Caterpillar' occurs in the word cluster with the words 'pollen' and 'Monarch' whereas the word 'larvae' is oriented towards a separate cluster with words like 'laboratory' and 'report,' that is, when referring to the research process. The science communication induces this distinction between the scientific word and the more common word usage.

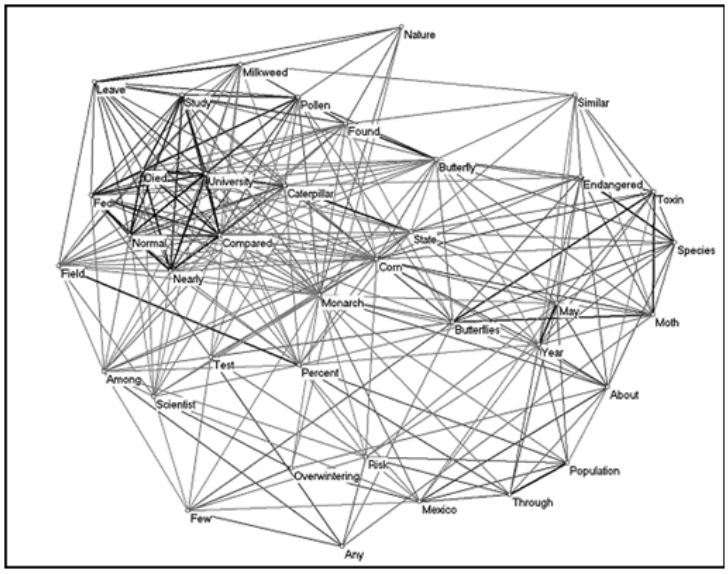


Figure 3: The cosine map of 38 words (cosine ≥ 0 . 5) .

The researcher expected that in the press release by the Union of Concerned Scientists (Figure 3), the words 'pollen' and 'Monarch' might again be presented in separate word clusters because this press release built directly upon the original letter in Nature. However, this was not the case. The UCS press release contains 7 paragraphs and 454 words. Only 38 words occurred more than once, and therefore form the basis for the semantic map. In Figure 3, the words 'Monarch' and 'pollen' appear as parts of the same component, although a bit more separated than in the university press release. In this document, the word 'Monarch' holds a central position. The frame has thus shifted from the genetically modified 'pollen' (the cause) to the Monarch butterfly as an endangered species (the consequence). The word 'larvae' is not used, and the term 'caterpillar' is part of the same word cluster as the words 'pollen' and 'Monarch.' The argument is mainly popularized.

The researcher selected the 38 words that occurred more than once for the visualization (Figure 4).

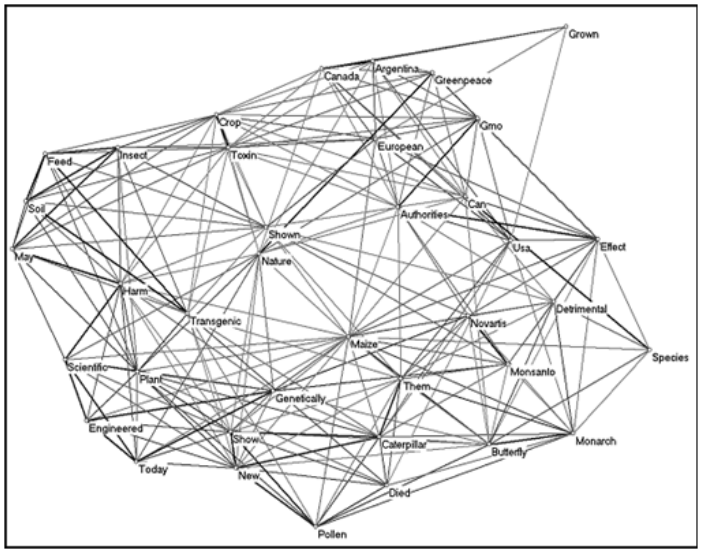


Figure 4: The cosine map of 38 words used more than once (cosine ≥ 0 . 5) .

In Figure 4, the words 'pollen' and 'Monarch' again belong to the same word cluster. The cluster is at the margin of the figure because the main concern is not with the discovery, but with its social consequences. Further, the word 'caterpillar' is part of the same cluster of words including 'pollen', 'died', and 'Monarch,'. However, the word 'larvae' is not used. The words 'Nature' and 'maize' hold central positions in the map.

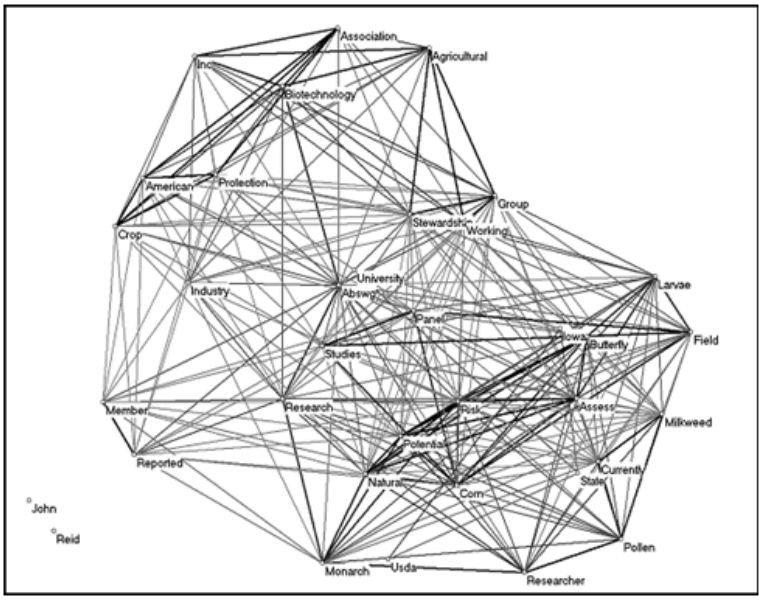


Figure 5: The cosine map of 36 words used more than once (cosine ≥ 0 . 5) .

The words 'pollen' and 'Monarch' are also part of the same word cluster (Figure 5). The document consists of 6 paragraphs that contain 361 words. Thirty-six words occurred more than once and were selected for the analysis. The word clusters are different from the map of Greenpeace in that the emphasis is on the 'potential risks' instead of scientific research that 'shows' the risks.

In conclusion, the researcher was able automatically to filter out semantic differences between these five documents. This could be analysed and visualized using the network of co-occurring words. However, our analysis remains purely semantic. One cannot indicate the rhetorical value of the claims without reading the documents, or without content analysis, because these pragmatic elements belong to another dimension of the communication.

The technique enabled us to detect that the main change in the semantics of the co-words occurred when the topic moved from the scientific context of Nature to the various press releases, including the press release by the university. The expectation of audiences seems to guide the selection of the frames of reference. In the semantic maps, one can also see novel topics across the various domains, such as the focus by the UCS on the butterfly instead of the pollen. While these five documents can also be coded manually, our purpose was to develop these techniques for larger document sets; the following two case studies use large sets of texts as data.

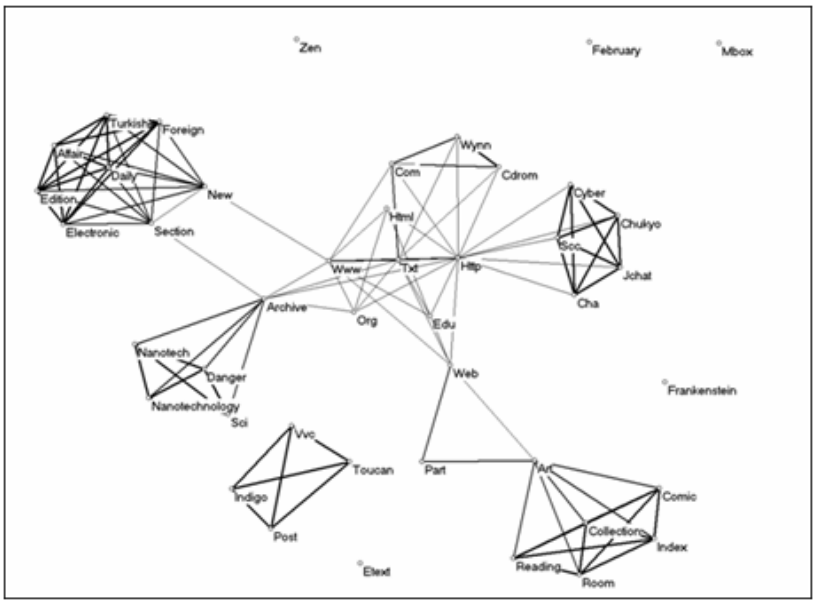


Figure 6: The cosine map of 44 words used more than once (cosine ≥ 0 . 1) .

In this semantic map (Figure 6), there are a few clusters of words that reflect the debate in discussion forums and archives on the Web. In the titles of the documents, the metaphor of Frankenfood was not yet used in 1996, and even the word 'Frankenstein' is still unrelated to the word clusters. Frankenstein food was an emerging topic in the AltaVista domain of that year. As the researcher used a list of stop words provided by the U.S. Patent Database for reasons of consistency, some of the most commonly co-occurring words on at the web like http, www, org, and edu were not suppressed. These words play a central role in the map in this relatively small set of title words. The other main clusters of words are around the dangers of nanotechnology, and news clipping. However, the metaphor was not yet established on the Internet at that time.

Figure 7: The cosine map of 100 words used more than 31 times (cosine ≥ 0 . 1) .

The interpretation of these results is as follows: the decline of the organizing power of the metaphor was rapid in 1999 and 2000 when the metaphors of 'Frankenfood' and 'Frankenstein food' began to be outdated. Due to its generalized meaning, the metaphor was used increasingly across domains and therefore lost its domain-specificity and the ability to organize distinctions among domains. This might also explain why the NGOs stopped using the metaphor in 2000 .From this perspective the metaphor can be considered as an anti-codifier: the metaphor mediates meaning among contexts and thus blurs boundaries. The three figures presented above show the life cycle of the metaphor. The year 1999 provides the peak in the codification among co-occurring title words. Further research is needed to see whether other kinds of metaphors function similarly, that is, whether metaphors function as anti-codifiers over time in being used across boundaries.

The number of documents that form the basis of the analysis seems to affect the results: the more documents analysed, the more variation the semantic maps show. In other words, a single document is more codified than a set of documents. In the next section, the researcher now proceeds from a single set of texts to a set of sets of texts, and explores how the differences among them affect the relative codification in the meanings of the words.

Measuring the meanings of 'stem cells' across domains

"Stem cells" have been an object of research since the 1960s. Progress in stem-cell research has been rapid from since 1990 and since the mid-1990s has provoked vivid public debate on the technical aspects involved. The 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 the term differently in the relevant domains.

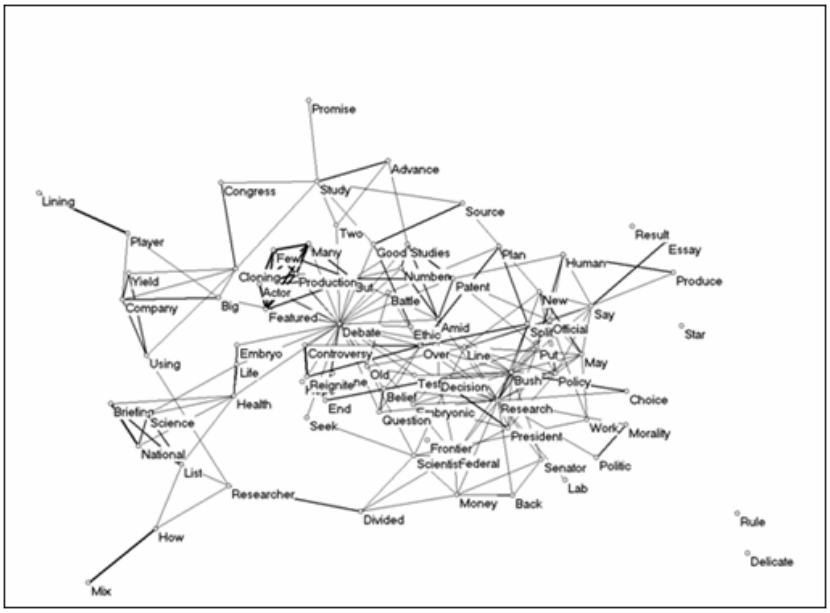


Figure 10: The cosine map of 81 title words used more than once (cosine ≥ 0 . 1) .

Inspection of Figure 10 shows us that the debate in the newspaper focused on the political agenda. The word 'debate' has the central position of a star in the network. One main cluster of words shows a representation of Bush's position, with words such as 'President,' 'Bush,' 'official,' 'policy,' and 'decision,' and on the other side the various aspects of the topic are reflected in words like 'embryo,'

'life,' 'health,' and 'science.' Some words (e.g., 'potential') that held a central position in Bush's speech, are not among the words that play a role in the structure of communication in the newspaper. Instead, the popularization of the issue for wider audiences draws from a wide variety of other relevant topics such as cloning.

In the newspaper, the words "stem cells" function as a metaphor that provides a reference to one of the debates on the national policy agenda. The specificity of word usage in this dataset is lower than in Bush's argument itself. As in the case of the press releases about Nature article on Monarch butterflies, the reflection reduces the codification. In other words, the word usage becomes more metaphorical.

As a third set of texts, the researcher analysed scholarly articles indexed in the Social Sciences Citation Index in 2001 with the words "stem cells" in their titles. The semantic map is based on the 41 words that occurred more than once in the titles of 53 documents (Figure 11).

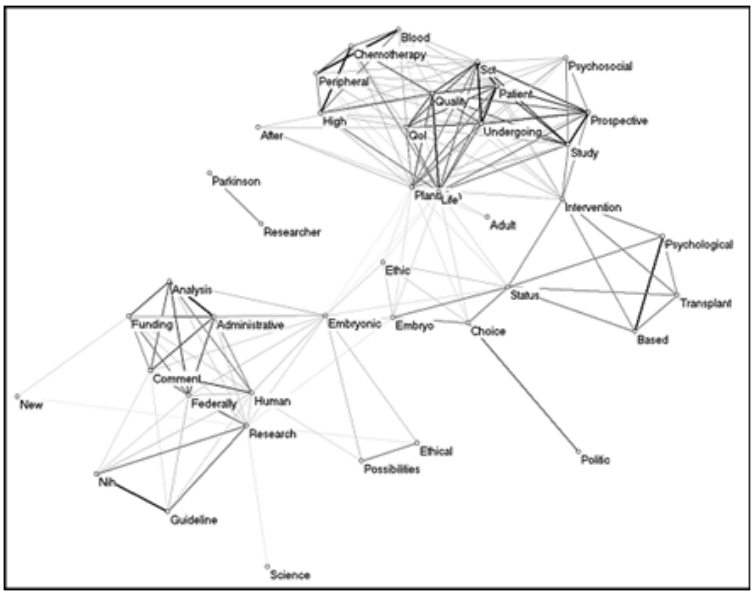


Figure 11: The cosine map of 41 title words used more than once in the 53 documents on stem cell in the Social Science Citation Index 2001 (cosine ≥ 0 . 1) .

In Figure 11, scholarly articles are differently codified into discourses: medicine, effects on patients, administration science (regulation), and ethics are all represented in the map. Specific words, such as 'status,' 'embryonic,' and 'intervention,' tie some of these clusters together. The different paradigms in these sciences operate as different codifiers. In other words, the words "stem cells" have a specific meaning in these different discourses, which counter-act acts upon the metaphorical function of these words in the public domain. Thus, it is well observed how the words "stem cell" can function as a metaphor in one context and as a diaphor in another.

In conclusion, the techniques presented here allowed us to automatically map the different degrees of codification of the words "stem cell" across the various places in the continuum between the sciences and society. However, this continuum is highly structured by interfaces.

FINDINGS

However, research shows that the meaning of new words encountered during reading is in fact

rarely guessable with complete accuracy, and proponents of the 'output hypothesis' emphasize the potential for learning through writing. Where dictionary training in some form is adopted, the weakest form is perhaps simply for the instructor to explicitly allow the use of dictionaries in classroom and homework tasks, rather than forbidding it and driving their use underground, leading possibly to a reliance upon poor dictionaries. Slightly stronger is for the teacher to evaluate what is available, recommend suitable dictionaries, and require their use in certain tasks. Dictionary evaluation (whether by teacher or student) is not easy given the large number of dictionaries often available, the unreliability of publisher's hype, and the fact that hardly any are ever subjected to independent research studies on their effectiveness. Typically a checklist approach has to be adopted which can usefully consider the various types of dictionary found appropriate to give users, their adoption of modern design features, and their suitability to the uses they need to make of them. If no more is done than the above, dictionary training is simply a version of the 'practice makes perfect' view of training, which may not be sufficient. To go further, the instructor may indulge in focused teaching of dictionary skills/use/strategies in some way.

(I)The more traditional approach presupposes comprehensive analyses of the relevant specific skills so as to create a syllabus. They are then taught via a regime such as: define it, give an example, e.g. by the teacher modeling its use, requiring the students to perform a task using the targeted skill. Workbooks exist associated with many dictionaries but though they often make the learner aware of a wide range of types of information offered by a dictionary they do not always do so in a way that trains the learner in the skills needed to access and exploit that information in real tasks.
(II)The more learner-centered approach shares much with humanistic pedagogy, and adopts a reflective approach to training. It is used in wider strategy training but rarely for dictionary use specifically. The instructor's role is to elicit from the trainees their own ideas about what they do and promote sharing and self-discovery through means such as: requiring students to keep a diary of their lookups (reasons, failures etc.), eliciting and sharing among a class their memories of their habits or experiences of dictionary use, and having them do think aloud tasks in pairs where they perform lookups.

SUGGESTIONS

This project is to develop tools that will help writers by showing them some of the alternative ways by which they can express their ideas.

From Proofing Tools to Writing Assistance

For all the huge leaps in progress in computing technology over the last half century, computers continue to be used extensively for one very old-fashioned purpose: creating text. Yet the range of tools aimed at helping writers with the authoring process has remained fairly static, with spelling and grammar checkers aimed at helping users avoid small or potentially embarrassing errors. Much less effort has been devoted to building tools or applications that assist writers in constructing better prose or finding alternative ways of expressing what they wish to communicate, in part because these have been seen as involving deep natural language understanding and therefore an almost intractable problem.

Helping writers find the right words

The present goals are more modest, and the research prototype offers a new spin on an old

technology: the thesaurus. Writers often have trouble coming up with just the right word to use in a particular context, or they may seek a little variety of expression, or they may need to follow the terminological conventions of a field or industry. For some of these purposes, a thesaurus can be of help, but the results are often not especially relevant in the context intended. For common words, the list of suggestions can be very long and esoteric, and yet somehow the right word never seems to be in the list.

The solution crucially involves:

•An enormous thesaurus containing 1 million keywords and key phrases. This resource dwarfs the typical desktop thesaurus, which might contain 300K headwords, and its size makes it much more likely that we'll find an interesting rewrite for any given word or phrase.

•English synonyms and phrasal paraphrases, e.g. express permission/explicit authority, learned as a byproduct of our group's data-driven Machine Translation effort. When we learn that two English words or phrases translate identically into another language, we can also infer that they might be similar in meaning in the right context.

•Very large language models that use sentence context to rank and filter thesaurus candidates in the same way that Word 2007's "Contextual Speller" uses context to decide which spelling variant (e.g. "you're" vs. "your") is most appropriate in a given sentence.

The result of all this is a new kind of thesaurus; one that does not simply point the user to a list of synonyms for a word in their document – most of them not quite right for one reason or another – but that instead suggests a smaller set of synonyms that are most likely to make sense in that particular context. The tool can even attempt to rewrite an entire sentence, selecting among different combinations of word and phrase replacements to choose the contextually most plausible set of all substitutions proposed by the models.

Some of the suggestions in this research are things that a writer might actually want to consider. The result is certainly far more usable than, say, random substitution of synonyms without reference to context, which produces delights.

It is also obvious that our process is far still from perfect - we would certainly advise against blindly adopting all suggestions that are offered. Errors do creep in when we've learned a bad English-English "translation" from our parallel translation data, or when the statistical models lack rich enough information to make the right decision about which alternative is most contextually appropriate.

In suggesting that writers replace content words, the researcher thus takes a great deal of risk: a poor choice can dramatically alter the meaning of a sentence - or provoke unintended hilarity.

We must take that risk, however, in order to push the frontiers of editing tool technology and the broader ability to identify and generate paraphrases. From a technical standpoint, the task of filtering potentially huge sets of synonymous words and phrases is itself immensely challenging. Our prototype is necessarily implemented as a web service, since the contextual language models required to make subtle judgments are so large.

The project goal is not to improve the work of poets, professional novelists, or anyone else who considers their writing art. The researcher focused primarily on helping users who are writing to achieve a more pragmatic goal – say a project report, a term paper, or an email – and who would like a little assistance in order to find the right words.

The long-term vision: learning to paraphrase

A common complaint about thesauri is that even when one of the suggestions is on-topic, it's only useful if the entire sentence is rephrased; the synonym cannot simply be plugged into the same slot as the original word. Currently, the tool suffers from this same limitation: we can only replace words or phrases in situ. The longer-term goals are loftier, and in particular, functionality that goes beyond simple word and phrase replacements and will offer more dramatic rewrites along the lines encountered in translating from one language to another, with wholesale rearrangements of words and phrases.

As we progress with this editing work, we anticipate borrowing more and more technology from our group's extensive work on machine translation. Paraphrasing one English sentence as another is essentially the monolingual version of translating from one language to another. Consider the following two sentences: the words and their order are quite different, yet at some level they "mean the same thing":

Recognizing and generating such paraphrase relationships is key to developing software applications that appear to "understand" natural language, since the same command, question, or fact can be expressed in myriad different ways. Rewriting prose in the context of a word processor is an application that interests us not only because users deserve better tools in this space, but also because it pushes this broader research agenda.

When attempting to decipher the meaning of a new word, it is often useful to look at what comes before and after that word. The surrounding words can give readers helpful context clues about the meaning and structure of the new word, as well as how it is used.

Using context clues aligns with the following ELA Common Core Standard:

Determine or clarify the meaning of unknown and multiple-meaning words and phrases by using context clues, analyzing meaningful word parts, and consulting general and specialized reference materials, as appropriate.

Helping struggling students use context clues

There are six common types of context clues , and teachers need to provide struggling students and those with learning disabilities with direct instruction in how to use these clues.

a. Root words and affix
b. Contrast
c. Logic
d. Definition
e. Example and Illustration
f. Grammar

Teachers have found it effective to model a self-questioning strategy to identify the different types of context clues. You can ask questions that are designed to focus attention on the unknown word and the possible clues to its meaning, such as: What are the surrounding words? How do these offer me clues? What does this word mean in terms of the context?

It is also helpful to provide students with frequent reminders and examples of the different types of context clues. One can also display the list on the bulletin board in your classroom so that students can easily remind themselves about context clues. Students can also keep examples in their

reading or writing portfolios.

Embedded supports in digital text

When students are reading digital texts, they can mark the text in a number of different ways to identify context clues. They can highlight, underline, bold, or vary the font (size, style, or color) of unknown words and/or the surrounding context. They can mark the clues that they believe will help them uncover a word's meaning.

Many supports can also be embedded in the digital text to help students while they are reading. For example, selected words and phrases (the unknown words as well as the surrounding context) can be linked to definitions, synonyms, antonyms, images, and audio explanations. The video, "Embedded Supports to differentiate Instruction for Struggling Students"" provides valuable ideas that can help you use embedded supports to differentiate instruction.

In addition to taking advantage of digital text, students and teachers may find it helpful to use an online dictionary and/or thesaurus. Visual Thesaurus is a dynamic "web" of words that can be expanded and reorganized by students.

In the classroom

He knows that while most of his students will understand the text, his struggling readers will need differentiated support to succeed. The specific objective of the lesson is to have his students review the different types of context clue, and to practice using them to define vocabulary words. This aligns closely with two specific ELA Common Core State Standards:

His students are able to read the digital text using a variety of devices, including tablets, e-readers, and laptops. He plans to introduce Visual Thesaurus to his students so they can practice using the relationships between words to help them define words that are unfamiliar. Some students will also need to use online dictionaries to confirm that their definitions are correct.

Lesson plan

Before Reading

- Review the skill: context clues.
- Explain the purpose of the lesson, building on past lessons using context clues.
- Model how to find context clues.

During Reading

- Display a short passage about on the interactive whiteboard.
- Have students read the passage in pairs.
- Invite students to come up to the whiteboard and highlight unfamiliar words and surrounding context clues.
- Have students share their thinking and discuss.
- To confirm their thinking, have students search for words in dictionaries and other resources.
- Repeat the process for one or two more words.
- Have students use context clues to understand new vocabulary and track these words on the class wiki.

After Reading

- Have students share their answers and prompt them to explain their process.

- In pairs, have students discuss at least two ways to figure out the meaning of a new word.
- Create a reference list of strategies.

DISCUSSION

In order to study semantic differences among individual texts and document sets, we have applied automated analysis of (co-)words in contexts to three different case studies. First, we analysed semantic differences in the frames of five documents because this allowed us to compare our results with those of an independent and previous content analysis. Second, we followed semantic changes over time in the structural dynamics of the co-word networks of Frankenstein foods. Third, we mapped semantic differences across various domains relevant for the debate on stem cells. In all of these case studies, we were able to map word meanings of the words independently from any a priori definition in a scheme or code book by taking into account both the relations of the words and the positions of these words in the distribution of relations. The researcher specified ex ante only the three flagship words of the scientific controversies.

The approach differs from that suggested by Callon et al. (1991), Ruiz-Baños et al. (1999) and Stegmann & Grohmann (2003) because these authors analysed co-occurrences amongst a set of key words. Their constructivist focus is on comparing the strength of the links, while our focus is on the structure in the constructed system of communication; specifically, how the words are positioned as a result of the linking and non-linking among them. Furthermore, this research is able to overcome some of the problems of the co-word analysis in the sociology of translation: first, our method is not limited to the key words assigned to the text documents; second, our technique can be applied to large sets without reducing the information content to the symmetrical co-occurrence matrix. The variables of the asymmetrical matrix of documents versus words can be considered as word vectors and accordingly we can use the vector-space model (Salton & McGill, 1983), while the normalization of the co-occurrence matrix has remained debatable. Third, by focusing not only on the relations between the words but also on their positions, we are able to measure the meanings of the co-words in their specific contexts.

In the debate on the effects of GM-pollen on Monarch butterflies, the researcher is able automatically to filter out some of the semantic differences constituting the frames of reference distinguished by Nucci (2004) on the basis of a content analysis. The semantic maps showed additional topics used in these domains, i.e., they demonstrated the structures in the contexts of communication. However, the semantic analysis could not inform us about the arguments made in these documents because the arguments belong to the pragmatic dimension of the communication. An analyst may have to focus on certain aspects in the semantic maps before the maps become meaningful.

In the case study of "Frankenfoods," we showed the dynamics of the network of co-occurring words over time. These networks changed from an emerging topic in 1996, headed for a clearly delineated and highly structured network in 1999, to a dispersed network of words in 2003. The metaphor of Frankenfoods functioned as an anti-codifier which blurs codified distinctions among domains over time. Further research is needed to specify whether other metaphors function similarly. Finally, in the case study focusing on "stem cells" we were able to show how the scientific and public contexts operate differently. The degree of codification is dependent on the context: a single text document is carefully constructed—therefore dense in its relations—and highly codified; a set of documents can be less codified and less densely packed. In the case of the Social Sciences Citation Index, however, the further differentiation according to disciplinary boundaries provided another structure. The word structure is highly organized by the scholarly reflection.

In this study, the researcher used pragmatic cut-off levels of approximately one hundred words for the semantic maps. A threshold was set in the case of 2003 AltaVista data because of the huge number of documents retrieved and the limits to visualization on a screen and is aware that this introduces error as did various other decisions, such as using a standardized stop word list across domains, etc. It is technically possible to include large numbers of words in the analysis.

The main argument is at another level and the researcher wished to show that the position of words in semantic fields can be used as indicators of their meaning. So, relatively straightforward standardizations of techniques in order are used, not to load the article with methodological details. The two concepts of positions and relations are associated because the relations add up and interact in a non-linear way; the positions are generated and stabilized within networks of relations. The analysis of relations between positions, however, requires the specification of a reflexive mechanism.

The results are based on normalizing the number of words included in the analysis without paying attention to the relative weights of the sets in terms of the number of documents or paragraphs within each unit of analysis. There is need for further research into normalizing the numbers of the units of analyses.

CONCLUSIONS

In conclusion, techniques for mapping the semantic meanings of co-words in contexts are suitable for automated filtering of the meanings of the words in their different domains of use, over time as well as across varying sets of texts. Focusing on specific functions of words—such as metaphors and diaphors—enabled us to specify the context in which these words gain their meanings. This specification enabled us to make the differences in the frames visible, to follow the development of codes of communication over time, and to analyze different degrees of codification used by various sides at the science/society interfaces. Hence, the method can be applied to a wide variety of longitudinal studies of science communication as well as comparative studies across the various domains of communication among the sciences and at science/technology/society interfaces .The differences among the domains of use inform us about the variation in the discourses, and about the selections in their respective operations. The methodology can also be used as an alternative to content analysis in the case of large (e.g., electronic) datasets that can no longer be coded manually. The study contributes to several research traditions that aim to automate the mapping of the dynamics of communications. On the one hand, the researcher is able to operationalize the mapping of the dynamics of knowledge. On the other hand, the specification of the context in which the co-words occur takes part within the debates on the sociology of translation and automated content analysis.

In the case studies the researcher has projected two reflexive mechanisms that are identified for the function of translation: metaphors and diaphors. The distributions are spatially arranged in networks. These networks are interfaced at each moment of time, but they contain codes which develop over time. Thus, there is both a dynamic and a synchronic aspect to the contexts. The operation of structures at each moment in time and their stabilization over time can be expected to lead to the globalisation or the decay of the knowledge base of codifications, due to the meta-stabilities that can be expected in the interactions among the differently codified sub dynamics of the communication.

REFERENCES

1.Ahlgren, P., B. Jarneving, & 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.

2.Bailón-Moreno, R., E. Jurado-Almeda, R. Ruiz-Baños, & J.-P. Courtial. (2005). Analysis of the Field of Physical Chemistry of Surfactants with the United Scientometric Model: Fit of Relational and Activity Indicators. Scientometrics, 63(2), 259-276.
3.Bernstein, B. (1971), Class, Codes and Control, Vol. 1: Theoretical Studies in the Sociology of Language. London: Routledge & Kegan Paul.
4.Burt, R. S. (1982), Toward a Structural Theory of Action. New York, etc.: Academic Press.
5.Callon, M, J. P. Courtial & F. Laville (1991), Co-Word Analysis as a Tool for Describing the Network of Interactions between Basic and Technological research: The Case of Polymer Chemistry.Scientometrics 22(1), 155-205.
6.Callon, M., J. Law, & A. Rip (Eds.). (1986), Mapping the Dynamics of Science and Technology. London: Macmillan.
7.Callon, M., J.-P. Courtial, W. A. Turner, & S. Bauin. (1983), From Translations to Problematic Networks: An Introduction to Co-Word Analysis. Social Science Information 22: 191-235.
8.Condit, C., B. Bates, R. Galloway, S. Brown Givens, C. Haynie, J. Jordan, G. Stables & H. Marshall West (2002), Recipes or Blueprints for Our Genes? How Contexts Selectively Activate the Multiple Meanings of Metaphors. Quarterly Journal of Speech 88(3), 303-325.
9.Fry, J. (2006). Scholarly Research and Information Practices: A Domain Analytic Approach. Information Processing and Management, 42(1), 299-316.
10.Hellsten, I. (2003), Focus on Metaphors: The case of 'Frankenfood' on the web. Journal of Computer-Mediated Communication 8(4) July
11.Leydesdorff, L. (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.
12.Losey, J. E., L. S. Rayor, M. E. Carter (1999). 'Transgenic Pollen harms monarch larvae.' Nature 399:214, 20 May 1999.
13.Maasen, S. & P. Weingart (1995), Metaphors -- Messengers of Meaning. A Contribution to an Evolutionary Sociology of Science. Science Communication 17 (1): 9-31.
14.MacKay, D. M. (1969), Information, Mechanism and Meaning. Cambridge and London: MIT Press.
15.Osgood, C. E., G. Suci, & P. Tannenbaum (1957), The Measurement of Meaning. Urbana: University of Illinois Press.
16.Stegmann, J. & Grohmann, G. (2003), Hypothesis generation guided by co-word clustering. Scientometrics 56(1), 111-135.
17.Weelwright, P. (1962), Metaphor and Reality. Bloomington: Indiana University Press.

Publish Research Article
International Level Multidisciplinary Research Journal
For All Subjects

Dear Sir/Mam,

We invite unpublished Research Paper,Summary of Research Project,Theses,Books and Book Review for publication,you will be pleased to know that our journals are

Associated and Indexed,India

- International Scientific Journal Consortium
- OPEN J-GATE

Associated and Indexed,USA

- EBSCO
- Index Copernicus
- Publication Index
- Academic Journal Database
- Contemporary Research Index
- Academic Paper Databse
- Digital Journals Database
- Current Index to Scholarly Journals
- Elite Scientific Journal Archive
- Directory Of Academic Resources
- Scholar Journal Index
- Recent Science Index
- Scientific Resources Database
- Directory Of Research Journal Indexing

Golden Research Thoughts
258/34 Raviwar Peth Solapur-413005,Maharashtra
Contact-9595359435
E-Mail-ayisrj@yahoo.in/ayisrj2011@gmail.com
Website : www.aygrt.isrj.org