UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

www.um.es/engphil/ijes

The Meaning of Genetics

SVENJA ADOLPHS, CRAIG HAMILTON. & BRIGITTE NERLICH
University of *Nottingham**

ABSTRACT
Research into the public understanding of genetics has greatly expanded lately. At the same time inatters relating to biotechnology have scizcd the public's attention. Corpus linguistics has long asked questions about how meaning is created and changed in the public sphere through language use. However, linking corpus linguistics to the study of the public understanding of science is something too few have done. To correct this trend, we apply methods from corpiis lingiiistics and cognitive linguistics to study how people talk about genetics. We do so by analysiny the meaning of words like *gene, genes*, *genetic*, *genetics,* and *genetically* as found in various spoken and written corpora. Specifically, we examine how they take on certain (e.g. figiirative) connotations and modulate in context.

KEYWORDS: Corpus lingiiistics, cognitive linguistics, gene talk, meaning change, context

I. INTRODUCTION

Access to multi-million word spoken and written corpora along with the development of sophisticated software tools to facilitate linguistic analysis has revolutionised language description over the past two decades. The description of word meaning through the analysis

* *Address for correspondence:* Dr. Svenja Adolphs. Dr. Craig Hamilton, and Dr. Brigitte Nerlich. School of English Studies and Institute for the Study of Genetics, Biorisk, and Society. University of Nottingham, Nottingham NG7 2RD, United Kingdom. Contact author: svenja.adolphs@nottingham.ac.uk

of concordance lines is an area that has developed most rapidly with the advance of computing resources and corpus evidence. While this type of methodology has become common practice in the field of lexicography, it has more recently been used in areas such as critical discourse analysis (Fairclough, 2000) and in the study of language and ideology (Stubbs, 1996). The advantage of using this technique in such contexts lies in the unmediated nature of corpus data which allows the analyst to tap into the way certain words are used in real-life contexts.

However, despite the interest in genetics in many different fields now, a precise analysis of what people mean by *genes* and other related words has yet to take place. Of course. as Lindsey (2001:3) argues, there is a difference between "gene talk" and "genetic communication". Whereas gene talk refers to the discussion of genes in lay contexts (i.e. contexts excluding scientists), genetic communication refers to the discussion of genes by professionals (i.e. medical geneticists, biologists and so on). Lindsey's point is that context can influence meaning, especially if various social groups have various different definitions about *genes* and other related words. It is for this reason that corpus linguistics, which can reveal in which context a word is used and how, offers a substantial bcnefit to linguists who study language in a variety of social contexts. But if corpus linguistics offers a sound method for a research project like this, one may ask why words like ***gene*** merit closer attention. We hold that gene talk is a fitting topic of analysis because it pervades our culture at present. After all, between 1953 and 2003, the fiftieth anniversary of discovery of DNA's 'double helix' structure (Dobzhansky, 1966; Leek, 1962; Osmundsen, 1961, 1964)[1], there have been tremendous changes in genetic science. Breakthroughs would include deciphering in the 1950s and 1960s what Watson and Crick called the human genetic 'code' (quoted in Nelkin, 2001:557) to cloning Dolly the sheep in 1997 or reporting the results of the Human Genome Project in 2000 and 2001 (Nerlich et al., 2002; Nerlich & Dingwall, in press). However, despite these breakthroughs, our relation to biotechnology seems ambivalent, fluctuating between hope and fear (see Smart, 2003, in press), and our language reflects this.

Our hypothesis is that the connotations associated with the leinma GENE will tcnd to be negative in the corpus data we study. We say this because public attitudes towards biotechnology, especially in Europe, are mostly negative (Mairis et al., 2001). Additionally, advances in biotechnology frequently receive sceptical treatment in the media (Bauer & Gaskell, 2002). Non-governmental organizations opposed to developinents in biotechnology, for example. havc been very successful in having the debatc framed conceptually on their terms rather than on the terms of the biotech industry (Hamilton, in press). In contrast, other genetic interventions intended to heal, cure, or prevent discase seem to be regarded as more positive developnients. Somewhere in between would be thc issue of the genetic modification of food although that too can swing either way[2]. In order to test our liypothesis, we aim to find out if one of the root causes for negative connotations can bc found in the seniantics of *gene, genes, genetic, genetics,* and *genetically* as those words are used differently by different people.

Our purpose is to highlight meaning pattems in the way that certain terms related to the field of genetics are used in context. In essence, there is no better laboratory for studying people's understanding of genetic science than looking directly at the relevant words themselves. In what follows, we first briefly discuss one of the methodologies developed in the area of corpus linguistics to describe the meaning of a lexical item and provide an overview of the corpora wc have chosen for our analysis. Then, in the analysis itself we concentrate on various parts of the lemma GENE. Finally, the results of our analysis are contextualised in the wider perspective of the emergence of meaning in various types of discourse. Where meanings are metaphorical, we argue that this is partly motivated by an underlying system of conceptual metaphors that structure human thinking and acting, and partly by the culture we livc in or the culture within which these meanings emerge (Zinken et al.. in press).

II. CORPUS LINGUISTICS AND THE UNIT OF MEANING

Recent advances in corpus linguistics havc highlighted the iniportance of syntagmatic relations in language use. Sinclair (1996) points out that it is difficult to ascribe meaning to individual words as strong patterns of co-occurrence with other words or classes of lexical items suggest that units of meaning are "largely phrasal" (1996:82). In order to describe the nature of individual units of meaning, Sinclair (1996) suggests four parameters: (1) colligation, (2) collocation. (3) semantic preference, and (4) semantic prosody. Colligation, the first parameter. describes the co-occurrence of grammatical choices. Grammatical patterning around a particular word accounts for the “variation” of a phrase, which "gives the plirase its essential flexibility, so that it can fit into the surrounding co-text" (Sinclair, 1996: 83). The notion of collocation, the second parameter, refers to the attraction between individual lexical items that regularly co-occur. For example, one of the main collocates of the adjectivc "genetic" is “engineering”.

There are a number of statistical procedures that can be used to account for lexical attraction. Such methods tend to compare thc expected frequency with which two words co-occur in a corpus with the actual frequcncy of co-occurrence. Two of the statistical measurements that have become common tools to calculate lexical attraction are the T-score and Mutual Inforniation'. Due to the limited space of this article we will not be able to discuss tliese measurenicnts in detail but we assumc that the higher these two scores are. the stronger thc indication that there exists a non-accidental relationship between the search word and its collocate.

The third of Sinclair's parameters is the "semantic preference" (Sinclair, 1996:86), a semantic abstraction of the prominent collocates of a lexical item or expression. Sinclair (1996:86) states: "This new criterion is another stage removed from the actual words in the text, just as colligation is one step more abstract than collocation. But it captures more of the patterning than the others". For example in his discussion of the expression "the nakcd eye", Sinclair finds that most of the verbs and adjectives preceding this expression show a scmantic preference of "vision". The verbs "see" and "seen" regularly occur in usage in the environment of "the naked eye".

Sinclair's (1996) fourth criterion in the description of the units of meaning is the "semantic prosody". In a discussion of semantic prosodies and irony. Louw (1993:157) argues that "evidence is emerging that departures in speech or writing from the expected profiles of semantic prosodies, if they are not intended as ironic, may mark the speaker's real attitude even where s/he is at pains to conceal it". Semantic prosodies, then, are associations with certain lexical connotations which are not easily detected by intuition. For example, Louw (1993:159) writes: "the habitual collocates of the form *set in* are capable of colouring it, so it can no longer be seen in isolation from its semantic prosody, which is established through the semantic consistency of its subjects". Louw's concept of "colouring". with regard to semantic prosodies, refers to what is seen as either a word's *negative* or *positive* semantic prosody. For instance, the word "happen" consistently takes a negative prosody (Sinclair, 1991). That is, bad things rather than good things appear to "happen". Stubbs (1995, 1996) and Sinclair (1991) both study lexical items that collocatc with negativc cvents, such as the word "cause" or "set in". Concordance searches of such items reveal that most of the nouns immediately following or preceding thcsc vcrbs are negative, such as "bad weather", "epidemic", etc. The fact that these items are recurrent in the concordance search is significant, as is the overall semantic field, or semantic preference they are related to.

Sinclair's criteria for describing the 'unit of meaning' of a lexical item then allow the analyst to include aspects of word meaning that help reveal conceptions about the area of genetics which were previously only open to speculation. In our analysis of the terminology associated with genetics we feel it is important to include a range of forms of this lemma as the individual items can show differences in the meaning profile as analysed within the framework outlined above. Based on the British National Corpus (BNC) we have identified the five most frequent items of this lemma: *gene, genes, genetic, genetically* and *genetics*. Our analysis is based on three corpora which are further described below.

III. DESCRIPTION OF CORPUS DATA

In order to gauge attitudes towards the terminology most closely related to the lexical item *gene,* we have chosen to study concordance output in three different corpora of contemporary spoken and written English: The Bank of English, the British National Corpus and the CANCODE corpus.

The Bank of English is a collaborative project between COBUILD and the University of Birmingham. It is one of the largest existing corpora to date and has mainly been used to inform dictionary design and lexicographical research. While more data continues to be added to this corpus. the latest release (January 2002) comprised 450 million words of spoken and written data. The majority of texts in this corpus originate from after 1990. The written component consists of texts from a variety of different sources including newspapers, fiction and non-fiction books, reports, letters, and magazines. The spoken part of the corpus is made up of everyday conversation, as well as radio broadcasts and a range of more formal spoken contexts, such as interviews and meetings.

The data for the British National Corpus (BNC) were collected in the early 1990s and it now consists of 100 million words of spoken and written British English[5]. The written part of the corpus accounts for 90% of the overall number of words and includes amongst other texts, newspaper cxtracts, journals, popular fiction. and academic books. The remaining 10% form the spoken part of the corpus and include informal conversation by a wide range of speakers, radio broadcast data and formal meetings.

The Cambridge and Nottingham Corpus of Discourse in English (CANCODE) is a collaborative project between the University of Nottingham and Cambridge University Press[6]. The main phase of data collection took place between 1994 and 1999 with a focus on gathering conversations from a variety of discourse contexts and speech genres. The 5 million word corpus consists exclusively of conversational data which were carefully selected to include adult speakers of different ages, sex, social backgrounds and levels of education[7]. The corpus itself has been organized according to five context types which represent a cline of formality. The framework of categorisation is based on the relationship that holds between the speakers in the dyadic and multi-party conversations in the corpus. These types of relationships fall into five broad categories which were identified at the outset and subsequently refined: *intimate, socio-cultural, professional, transactional,* and *pedagogic* These categories were found to be largely exclusive while being comprehensive at the same time. In the intimate category the distance between the speakers is at a minimum, such as is the case in interactions between partners or family members. The *socio-cultural* category implies the voluntary interaction between speakers that seek each other's company for the sake of the interaction itself. The relationship between the speakers is usually marked by friendship and is thus not as close as that between speakers in the *intimate* category. Typical venues for this type of interaction are social gatherings, birthday parties, sports clubs. and voluntary group meetings. The *professional* category refers to the relationship that holds between people who are interacting as part of their regular daily work. As such, this category only applies to interactions where all speakers are part of the *professional* context. The *transactional* category embraces interactions in which the speakers do not previously know one another. The purpose behind transactional conversations is usually related to a need on the part of the hearer or the speaker. As such, the conversations aim to satisfy a particular transactional goal, such as buying and selling for example. The *pedagogic* category was set up to include any conversation in which the relationship between the speakers was defined by the pedagogic context. A range of tutorials, seminars and lectures were included[8].

We will draw on all of the three corpora outlined above in our analysis. They combine to a useful sample of spoken and written British English used in the 1990s. The respective classification schemes that have been applied to these corpora allow the analyst to make some statements about context specific use of the lexical items under discussion.

IV. RESULTS AND DISCUSSION

The individual lexical items in the lemma GENE were chosen according to frequency criteria. An initial frequency count in the British National Corpus revealed the following figures:

Gene:	2237
Genes:	2069
Genetic:	1823
Genetically:	335
Genetics:	302

The nouns *gene* and *genes* are clearly the most frequent representations of this lemma. They are closely followed by the adjective *genetic*. The frequency drops sharply when we consider the adverb *genetically*. The noun *genetics* has the lowest frequency.

While the British National Corpus offers us a general picture of frequencies, we can turn to the CANCODE corpus to analyse frequencies according to different conversational contexts.

Figure 1: Frequencies of the lemma GENE as found in the CANCODE corpus

	intimate	socio-cultural	professional	transactional	pedagogic
gene	4	1		1	16
genes	13	1			35
genetic	6	5		2	44
genetically	3				1
genetics	2				3

These frequency results are interesting as they give us an indication of the types of situations in which people discuss *genetics*. While we would expect these figures to be high in the pedagogic category. taking into account that a number of recorded interactions were medical and biology lectures and seminars. it is interesting to note that the area of genetics is also being discussed in interactions between close friends and partners. It is exactly this type of social sphere where unmediated recorded conversations can offer us insights into public attitudes to genetics and we will return to this aspect below[9].

If we consider the concordance lines taken from the intimate and the pedagogic category, it becomes clear that the collocations are different between the two. In the intimate category the adjective *genetic* pre-modifies the nouns *mutation, programming* and *experiment* and thcre is some cvidence of a negative prosody in this sample (e.g. 'mutation', 'vile and foul genetic experiment'). The examples taken from the pedagogic category collocate with *material* and *pool,* and although they display a semantic prosody of negative events (e.g. *viruses, abnormalities* and *failure),* there seems to be no personal opinion included in them.

Intimate:

```
          $1> It could be a genetic mutation.
e same way that red hair was genetic programming which has skipped a
                 <$=> But a genetic experiment cloning a naturally c
f a mouse is vile and a foul genetic experiment.
just saying that it could be genetic programming.
 Well they come out the same genetic </$O8> they come out the same </
```

Pedagogic:

```
you and I carry D N A as our genetic material and it's double-strande
               Three K Bs of genetic material.
 the way viruses carry their genetic material.
        If we get a change in genetic pool this can lead to a failure
    Viruses consist of some genetic material whether it be R N A or
 abnormalities or changes in genetic pool.
```

We then considered the respective units of meaning of the chosen lexical items when we looked at the British National Corpus and the Bank of English. Here our analysis starts with the two most frequent items, the nouns *gene* and *genes*. Both often occur as modifiers in complex noun phrases or as part of compound nouns (e.g. gene pool, gene therapy, gene activity, snail genes, genes code, etc.). In this form they tend to be used as extended metaphors, an issue to which we will return later. The contexts in which both nouns are found are almost always scientific which is mirrored by the semantic preference of biomedical vocabulary as shown in the examples taken from the BNC below:


```
                                    Gene 62, encoding the 140k protein, lies
   coupling a promoter region of a gene expressed in cancer cells with a
            Screening of the EMBL3 gene bank with terminal Eco RI fragments
     controlled by one or even two gene patterns in the DNA

   originated as 'own' chromosomal genes.
by the function of normal cellular genes.
     of alga genes and Chlorohydra genes coincide
       the chromosome which has no genes and by which the chrornosome
```

The same tendency emerges from an analysis of thc 10 most significant collocates of the lexical items under discussion. The results in the table below are based on a sub-sample of the Bank of English which is available for demonstration searches on-line. All results are based on T-score analyses.

Figure 2: Ten Most Significant Lexical Collocates for the Lemma *GENE* based on the Cobuild Direct on-line collocation sampler

Collocates	gene	**genes**	**genetic**	**genetically**	**genetics**
1	hackman	human	engineering	engineered	biology
2	therapy	other	material	modified	research
3	kelly	cells	defects	programmed	human
4	responsible	cell	research	been	biochemistry
5	sarazen	specific	disease	have	molecular
6	gene	disease	make	determined	behaviour
7	cancer	language	DNA	food	cancer
8	scientists	cause	factors	different	role
9	disease	inherit	environmental	foods	microbiology
10	called	inserted	differences	cells	genetics

However, while collocations and semantic preferences of these words tend to merely give an indication of the general contexts in which items of this lemma are used, certain aspects of the negative semantic prosody are already becoming apparent at this stage. This aspect is realised in collocates such as *cuncer* and *disease*. It should also be noted that one of the collocates of the word *genes* in the table above is the lexical item *cause* which, as Stubbs (1995) demonstrates. has a strongly negative semantic prosody, hinting at the pervasive popular belief that genes 'cause' disease or behaviour, a mistaken belief commonly referred to as ‘genetic determinism' (Dennett, 2003).

A closer look at the concordance output reveals the general trend towards this prosody. It is interesting to note, however, that this trend is much more prominent with the lexical items *genetic* and *genetically* compared to *gene*, *genes* and *genetics* which display a more neutral, scientific semantic prosody. The concordance lines below which are taken from the BNC illustrate this:


```
concept of mutation which was a random genetic change
  examples  of these are the inherited genetic diseases such as
  female was therefore a solution to a genetic conflict
                             the shadow of genetic injustice
                      tumours, accidents, genetic problems and meningitis can

become handicapped not because it is genetically abnormal or have a
  will organice consumer boycotts of genetically engineered foods
apply to morality issues surrounding genetically engineered organisms
                        not yet fat but genetically programmed for early
                    transmitted between genetically susceptible people
```

We can see from these concordance lines a general negative semantic prosody which emerges from surrounding vocabulary such as *canflict*, *injustice* and *abnormalities*. If we consider the nouns *gene*, *genes* and *genetics* on the other hand, the contexts remain more neutral overall as illustrated in the following concordance lines:

```
controlled by one or even two gene patterns in the DNA
Two related zinc finger (ZNF) gene clusters from the pericentromeric
            Expansions of the gene 62 promoter sequences

  powerful way of studying how genes affect development
  electronic database of human genea, so doctors around the world
that they have gets two normal genea, one from each parent

                         the principles of genetics.
learn about the structure, physiology, genetics and other properties
             and an enormous growth of genetics took place and indeed is
```

Space limitations mean we can only list here a very small sample of what reflects the overall patterns found in the concordance output most clearly. The concordance lines highlight the more neutral tone in the contexts where the items *gene*(s) and *genetics* are being used. Our data suggest the latter is used in more general contexts rather than in very specific medical or biological types of discourse.

Overall, we can say that the semantic prosody is closely related to the grammatical role of the particular lexical item of this lemma. The adjective *genetic* and the adverb *genetically* give rise to a range of processes that rcinforce the negative semantic prosody, such as *genetic modification* or *genetically engineered.* We would therefore argue that while there is a neutral, scientific voice to the discourse related to genetics in certain contexts, there are at the same time negative prosodies which are particularly prominent in the use of the lexical items *genetic* and *genetically*. The extract below illustrates this. It is taken from the

CANCODE corpus and shows a young couple discussing one of the aspects in the debate surrounding *genetics:*

<S01> I mean what he said was Yes I could agree with it. I'm not into altering natures like watching animals being...
<S02> Like *genetically* altered. Yeah.
<S01> Yeah tha= That I don't go with. Ifyou get bad ones then+
<S02> Yeah.
<S01> +you're meant to have bad ones+
<S02> Mm.
<S01> +in life. Tha's what makes life life isn't it? I mean it is sad when it's children but it is a, it's been like this forever, And you start altering that then that that is not on to me. But yet I don't want to be told+
<S02> Yeah.
<S01> +over, and about the crops and things. And he said. and he said about you know "Buy them in the shops these. Always look for organically grown as well". And I said to your dad "Everybody should be". And I said "Oh yeah. People with not much money are certainly gonna go and+
<S02> Yeah.
<S01> +look for organically grown stuff". "It only costs a few pence more". I thought "And the rest". You know stuff is so dear.
<S02> Yeah.

Here we see a representation of a negative attitude towards the process of *genetic altering* in a stretch of ongoing discourse which is conveyed in a series of statements of opinion (e.g. 'I'm not into altering natures', 'That I don't go with'). A good deal of research has been done on public opinion in Europe regarding genetics and biotechnology (Durant,et al. 1998; Wagner. et al. 2002), with men tending to favour genetic modification and biotechnology more than women, although attitudes across Europe can vary. In northern Europe, which includes the United Kingdom, people tend to have lower opinions about genetic and biotechnologies while in southern Europe, which includes France and Italy, people tend to have higher opinions of these technologies. The attitudes expressed above in the conversation thus reflect opinions common among Europeans today.

V. METAPHORS

We mentioned in our introduction some of the rapid breakthroughs secn over the last fifty years in genetic biotechnology. However, although new scientific developments may constantly change how we view technology, there is conceptual continuity as far as genes are concerned. This comes in the form of the metaphors used to describe DNA (and by extension, the human genome) for the last four decades or so. Here the corpus linguistic approach is complemented by a cognitive linguistic view of figurative language use. Some of the most pervasive and enduring metaphors for the human genome and DNA include the following conceptual domains (Pollack, 1994; Ridley, 2000)[10]:

DNA/THE HUMAN GENOME IS A LANGUAGE
DNA /THE HUMAN GENOME IS A CODE
DNNTHE HUMAN GENOME IS A MAP
DNNTHE HUMAN GENOME IS A TEXT
DNNTHE HUMAN GENOME IS A BOOK
DNNTHE HUMAN GENOME IS A BIBLE (HOLY BOOK)
DNA/THE HUMAN GENOME IS THE BOOK OF LIFE

Even the subcomponents of DNA have been conceptualised metaphorically, and this since the very beginnings of niodern genetics in the 1950s (Bygrave, 2002):

BASES ARE LETTERS
CODONS [FOUR-LETTER GROUPINGS] ARE WORDS
GENES ARE SENTENCES
CHROMOSOMES ARE CHAPTERS

For four decades these metaphors have been remarkably stable despite changes in genetic science. These textual source domains entail that to understand genes or DNA one must be able to read, since reading in this case is metaphorical for comprehension. To *read* the book of life is thus to understand the function and expression of genes in organisms. The reading metaphor first originated in the decision to name the four bases of DNA. Adenine, guanine, tliymine and cytosine were then represented by the *letters* A, G, T, and C. The fortuitous choice to represent bases by tlie first letters of their scientific names made it easier for metaphors to take hold in genetics that exploit our everyday knowledge of reading, books and codes. If the DNA bases had been represented by numbers, the whole metaphor system surrounding genetics might have looked different. Even the title of the 1997 film, *Gattaca*, was no doubt inspired by the letters used to stand for the four DNA bases.

However. it should be stressed that the reading metaphor also pre-dates the discovery of DNA. It ties in with the conceptual metaphor KNOWING IS SEEING (Sweetser, 1990:38) on the one hand, and has been used in westem culture to elevate the knowledge achieved by the natural sciences to the status of that represented in the holy book of the Bible (at least since Francis Bacon and Galileo Galilei) on the other hand. For physical vision to refer to mental "intellection", as Sweetser (1990: 38) suggests in her classic analysis of this pattern. we must map a physical domain onto a mental domain. For example, to say "I see what you mean" to indicate "I know what you mean" is just one of the many linguistic manifestations of the KNOWING IS SEEING conceptual metaphor. If to see DNA is to *know* DNA, then to read it physically is to understand it mentally. As Sweetser documents, the semantic shift from the physical to the mental that our perceptual verbs reveal offers solid evidence for the pervasive nature of this cognitive act in language. That is, polysemy is often motivated rather than arbitrary. There are good reasons. in other words, for using *see* when meaning *know*. With regard to genes, therefore, we fully understand why in June 2000 the then Vice President of the United States Al Gore compared the human genetic code to the Nazi secret code in relation to diseases like cancer (i.e. the enemy): "With the completion of the Human Genome, we are on the verge of cracking another enemy's secret code" (quoted in Annas 2000:775, note 79). Gore's statement implies that diseases hide in a secret code, which is bad, and that cracking that code may mean finding a cure for diseases, which is good. This is an assumption based fundamentally upon the KNOWING IS SEEING conceptual metaphor.

As for our hypothesis about negative semantic prosody for words akin to *genes*, there does seem to be ample support for this hypothesis from evidence in the metaphors in which these words are embedded. Consider, for instance, the following examples from the three corpora we have studied:

- genetic alchemy
- genetic control
- the DNA genetic control centre
- genetic engineering
- genetic manipulation
- genetic fingerprinting
- the shadow of genetic injustice
- burdened with their genetic lot
- risks of genetic pollution
- genetic discourse meets environmental discourse
- exact genetic replicas
- just an automaton driven by his genetic predisposition
- man-made genetic time bombs

As we noted earlier, the grammatical function of the adjective *genetic* seems to influence the negative semantic prosodies seen above. As Aristotle noted long ago, epithets (i.e. adjectives) could be metaphoric (*Rhetoric* 1405a:169), so we need not be surprised to find figurative phrases prompted by the adjective 'genetic' here. If genetically manipulated organisms are typically imagined as un-natural, artificially produced, robot-like creatures who evoke various stereotypes in literature and film, from Frankenstein to the Attack of the Clones in *Star Wars* (Nerlich, Dingwall & Clarke, 1999). then the metaphors serve a purpose.

As someone in one corpus put it, "Perhaps it was a genetic thing", whereby they may have meant to refer to something that caused a certain effect (e.g. cancer). If genes are that which cause life forms to exist, then understanding genes in causal terms makes sense. But as Sinclair (1991) discovered with his analysis of "happen", the term is generally used in such a way to provide it with a semantic prosody that is negative. In general, the same is true with words based on the lemma GENE.

The negative semantic prosody for the adjective *genetic* is also found for the adverb ***genetically.*** In the corpora, ***genetically*** seems to collocate with less-than-favourable words such as:

- genetically controlled
- genetically defined
- genetically determined
- genetically manipulated
- genetically mapped
- genetically modified
- genetically altered
- genetically based inferiority
- genetically handicapped
- genetically predetermined
- genetically prograrnmed
- genetically engineered

The clash between the natural and the artificial here could not be clearer. The "natural" in this case referring to something that is not controlled, manipulated, modified, altered, programmed, or engineered by human beings. All deviation from what is "natural" strikes us as dangerous, base, or something to be avoided. When we sense that something is out of our hands or that we can't do much about it, this implies that we are what we are because our genes determine who we are and so all human agency is removed from life. For these reasons we posit that there are negative connotations with the terms in the list above. Nobody will have any concern for a bridge that was "mechanically engineered" because of what we take the nature of engineering to be. However, when engineering meets biology, then concerns spring up because a tomato that is "genetically engineered" is perhaps something to reconsider before dinner.

The semantic prosody, therefore, reflects a limit of acceptability, especially since what ***genetically*** immediately precedes is not often a positive term in its own right. We accept the fact that cars are engineered, but we have a harder time accepting the fact that tomatoes too can also be engineered. From expecting things like cars, bridges, machines, computers, and so on to be humanly engineered, we have moved into a situation where we find that things which we did not expect to find engineered, such as viruses, food, crops, organisms, microbes, micro organisms, and farm animals, are in fact engineered in the way that a car is engineered today. This shift, from the natural to the artificial, points to the clash referred to earlier and relates directly to our understanding of control. As Nelkin (2001: 556) states, "Through metaphors, genetics can seem to be a source of salvation or a means of exploitation or control", although we should note that control can be seen in a negative and positive light. After all, it is perhaps desirable to 'control' the gene for Huntingdon's disease, for example.

There are, of course, a few exceptions from time to time where semantic prosody is concerned and there are cases where the semantic prosody may not appear to be negative:

- genetic endowment
- genetic richness
- genetic diversity
- genetically pure
- genetically purer
- genetically superior
- gene therapy
- gene pool
- genes might have been nature's gift
- desirable genes

These instances seem to suggest a positive prosody given the collocates with the lemma *GENE*. But, at closer inspection some have rather negative connotations in certain contexts. Genetic purity and genetic superiority are concepts that, in the context of eugenics, are unattractive. Some of these connotations may not be directly visible in the corpus because the corpora are less than 20 years old. But they come with the words as their ideological baggage from a time when genetic purity and genetic superiority were pursued by various states around the world, most memorably the Nazi state.

To understand in detail the source of our ambivalence towards genetic biotechnology, we need look no further than our very words: those based on the lemma *GENE* seem so often to be negative rather than positive. However, what Louw (1993: 159) would call "colouring" can occur in exceptional cases where the semantic prosody, say, for *genetic* is more positive than negative, as in *genetic richness*. And so, although the semantics tend to be more negative than positive, it is not impossible to get positive semantic prosodies when gene talk occurs. One reason for the occasional exception would have to be the rhetoric surrounding biotechnology. As Iina Hellsten (2002:5) remarks:

> The public debates on biotechnology and biodiversity are thoroughly rnetaphorised. Cloning is constantly discussed as if it dealt with the *mass production* of cornrnodities, either producing *lousy* ***copies*** *of the* ***original*** or *perfect products*. The Hurnan Genorne Project is expected to *reveal the secrets of life* but it is also opposed by wamings of *science playing God* —depending on the underlying views on the goal of this science's journey. Sirnilarly. the conscrvation of biodiversity is constantly discussed in terms of *the common heritage* of the hurnan kind, *richness* that should be preserved for future generations. This richness is sornetirnes defined as *gold, treasuries,* and *jewels* but sornetirnes also as values and *complex relations* —depending on the underlying views on 'nature' as either a store of cornrnodities or a dynamic network of processes.

In other words. there are good reasons for feeling ambivalent about what we are talking about when we are talking about genes. Those with a favourable view of genetic science might more frequently use positive meanings, whereas those with more doubts about the benefits of genetic science might more frequently use negative meanings. What our research shows, however. is that the champions of biotechnology have their work cut out for them given the fact that the words themselves used for genetic science tend more often than not to strike us a as negative rather than positive.

VI. CONCLUSION

As we have shown, a cognitive linguistic approach to semantics can be bolstered by the use of corpus linguistics. Our combined methods yield a new understanding of gene talk in all of its various aspects. That the semantic prosodies for words based on the lemma ***GENE*** are often negative in the corpora that we have studied highlights conceptual issues underlying current debates surrounding biotechnology. Moreover, the substantial use of metaphors in gene talk reinforces our view that metaphors are ubiquitous in everyday language. Given the complexity of genetic science and the invisible nature of genes, it is hard to talk about *genes* literally. This suggests that context counts where semantic prosodies are concerned. Meanings cannot be removed from pragmatic contexts (i.e. where gene talk occurs) or from lexical contexts (i.e. the words found before or after the word under analysis). As most of our data come from the 1990s we could only provide a synchronic snapshot of how the lemma ***GENE*** was used at a time when genetic breakthroughs revolutionised our understanding of life. and of the meaning of life, during the last decade or so. It would be interesting to analyse the diachronic changes in the uses and meanings of ***gene*** over time, from the 1960s, when genetic science had its first scientific and popular peak to the 1990s, when it had its second, but this will have to wait for another time. We hope to have shown, however, that combining methods from corpus and cognitive linguistics enables us to see what ***genetics*** means and why it means what it means when people talk about genes. As we found, to try to talk about genes in a positive manner is not easy. Those trying to sell biotechnology or genetic science might wish to keep this mind.

NOTES

For more examples, see Condit (1999).

For more infotmation on public attitudes towards GM food and designer babies, please see Turner (in prep.) and Townsend & Clarke (in prep.).

[3] See Stubbs (1995) for a discussion of using statistical analyses in the area of corpus linguistics.

[4] Please see Stubbs (1996) for a discussion of this issue.

[5] For a detailed description of this corpus, please see Burnard (1995) and Aston & Burnard (1998).

[6] The corpus was sponsored by Cambridge University Press with whom sole copyright resides.

[7] For a comprehensive description of the CANCODE corpus, please see McCarthy (1998).

[8] For a more detailed description of these categories, see Adolphs & Carter in this volume.

[9] It is important to highlight that the instances summarised in this table were drawn from a range of different conversations in the respective categories. We recognise, however, that our observations are limited to a small set of instances found in the corpus and that too much should not be claimed for them until further evidence is gathered.

[10] As so often with 'dead metaphors', the metaphorical roots and ramifications of such tetms only resurface in jokes, such as the one heard on 7 October 2001 on a popular BBC Radio 4 *News Quiz*. Somebody jokingly said on that programme that scientists had deciphered the genome of a plague virus which contained almost as many letters as the BBC's complaints department.

REFERENCES

Adolplis, S., & Carter, R. (2003). ***And she's like it's terrible, like:*** Spoken Discourse, Graininar aiid Corpus Analysis. *International Journal of English Studies* 3 (1): 45-56.

Annas, G. (2000). The Maii on tlie Mooii, Immorality, aiid Otlier Millennial Mytlis: The Prospects aiid Perils of Huinaii Geiietic Engineering. ***Emory Law Journal***, 49:3, 753-782.

Aristotle. (1954). ***Rhetoric and Poetics***. Traiislated by W. R. Roberts. New York: Raiidoin House.

Aston, G., & Buriiard, L. (1998). *The BNC Handbook.* Edinburgh: Edinburgli Uiiiversity Press.

Bauer, M.W. & Gaskell, G. (Eds.) (2002). *Biotechnology: the Making of a Global Controversy.* Cainbridge: Cainbridge Uiiiversity Press.

Buriiard, L. (1995). *British National Corpus: Users Reference Guide.* Oxford: Oxford Uiiiversity Coinputiiig Services.

Bygrave, M. (2002). Tlie Year of tlie Gene. *The Observe,:* 22 Deceinber: 21-22.

Coiidit, C. M. (1999). *The Meanings of the Gene: Public Debates about Human Heredity.* Londoii: Uiiiversity of Wiscoiisiii Press.

Deiiiiett, D. (2003). Tlie Mythical Tlireat of Geiietic Deteriniiiism. *The Chronicle of Higher Education Review*, 31 Jaiiuary: B7.

Dobzliaiisky, T. (1966). Tlie Code was Brokeii. *New York Times* 17 April: 3 (Sectioii VII).

Duraiit, J., Bauer, M. W. aiid Gaskell, G. (Eds.) (1998). *Biotechnology in the Public Sphere: a European Sourcebook.* Loiidoii: Scieiice Museuin.

Fairclough, N. (2000). *New Labotu-, New Language?* London: Routledge.

Hamilton, C. (in press). Geiietic Roulette: On tlie Cognitive Rhetoric of Biorisk. In R. Dirven (Ed.), *Cognitive Models in Language nnd Thought: Ideologies. Metaphors, and Meaning.* Berlin: Moutoii de Gruyter.

Hellsteii. I. (2002). *The Politics of Metaphor: Biotechnology and Biodiversity in the Media.* Acadeinic Dissertation, Acta Uiiiversitatis Tainpereiisis 876. Tainpere, Finland: University of Tainpere.

Leek, J. M. (1962). Tlie Code of Life. *New York Times*, 28 January: 8 (Section IV).

Liiidsey, N. (2001). *What is Gene Talk? Some Conceptual Issues.* Paper given at tlie Annual Postgraduate Foruiii oii Geiietics aiid Society, Uiiiversity of Nottingliam, Juiie.

Louw, B. (1993). Irony in tlie Text or Insincerity in tlie Writer? Tlie Diagnostic Potential of Seinantic Prosodies. In Barker, M. et al. (Eds.). *Text and Technology*, 157-176. Ainsterdatn: Benjamins.

Marris. C., Wyiiie, B., Siininoiis, P., & Weldon, S. (2001). *Public Perceptions of Agricultural Biotechnology in Europe.* Final Report of tlie PABE Researcli Project. Brussels: Commission of Europeaii Communities.

McCarthy, M. (1998). *Spoken Language and Applied Linguistics*. Cainbridge: Cainbridge Uiiiversity Press.

Nelkin, D. (2001). Molecular Metapliors: tlie Geiie in Popular Discourse. *Nature Reviews: Genetics*. 2, (July), 55-559.

Nerlicli. B., Clarke, D.D., & Diiigwall, R. (1999). Tlie Iiiflueiice of Popular Cultural Imagery oii Public Attitudes Towards Cloning. *Sociological Research Online* 4:3. Available online at <http://www.socresonline.org.uk/socresonline/4/3/Nerlich.html>

Nerlicli, B., Diiigwall, R.. & Clarke. D.D. (2002). Tlie Book of Life: How tlie Huinaii Geiioine was Revealed to tlie Public. *Health: An International Journal For The Social Study Of Health. Illness And Medicine,* 64, 445-469.

Nerlicli, B., & Diiigwall, R. (in press). Deciphering tlie Huinaii Geiioine: Tlie Seinaiitic aiid Ideological Fouiidatioiis of Geiietic aiid Geiioiiiic Discourse. In R. Dirveii aiid R. Fraiik (Eds.), *Cognitive Models in Language and Thought: Ideologies, Metaphors, and Meaning.* Berlin: Mouton de Gruyter.

Osinuiidseii, J. A. (1961). New Way to Read Life's Code Fouiid. *New York Times*, 7 April:12

Osinuiidseii, J. A. (1964). Breakiiig tlie Code. *New York Times,* 2 August: E7

Pollack, R. (1994). *Signs of Life: The Language and Meaning of DNA*. Bostoii & New York: Houglitoii Miffliii.

Ridley, M. (1 999). *Genome: The Autobiography ofa Species in 23 Chapters*. Loiidoii: Fourth Estate

Sinclair, J. McH. (1991). *Corpzo. Concoi-dance. C'ollocation.* Oxford: Oxford University Press.

Sinclair, J. McH. (1996). The Search for the Units of Meariiiig. *Textus,* IX: 75-106

Sinart. A. (2003, in press). Reportirig the Dawii of tlie Post-geiioinic Era: Who Wants to Live Forever? *Sociology of Health nnd Illness.* 25:1. 24-49.

Stubbs, M. (1995). Collocatioiis aiid Semantic profiles: oii tlie Cause of tlie Trouble with Quaiititative Metliods. *Functions of Language,* 2:1, 1-33.

Stubbs, M. (1996). *Text and Corpus Analysis*. Oxford:Blackwell.

Sweetser, E. (1990). *From Etymology to Pragmatics*. Cainbridge: Cainbridge Uiiiversity Press

Townseiid, E. & Clarke, D. D. (in prep.). *Risk, Feelings and Biotechnology: An Exploratory Study Using Message Framing.* Paper read at tlie BPS Social Psycliology Sectioii Coiifereiice. Septeinber 2002. Uiiiversity of Huddersfield.UK

Turner, S. (in prep.). Coiiflicts in tlie Perfect Babies Discourse. Special issue of *Discourse and Society* oii geiioinics discourse (Edited by Nerlicli, B., Dingwall, R., & Martiii, P.).

Wagner, Wolfgaiig, et al. (2002). Paiidora's Geiies—liiiages of Geiies aiid Nature.In Martiii Bauer aiid George Gaskell (Eds.), *Biotechnology. The Making of a Global Controversy.* Cainbridge: Caiiibridge UP. 244-278.

Ziiikeii, J., Hellsten, I., & Nerlich, B. (iii press). Wliat is Cultural about 'Coiiceptual Metaphor'. *International Journal of Communication.*