Mechanisms of Meaning

Patrick Hanks[(✉)]

University of Wolverhampton, Stafford Street, Wolverhampton, WV1 1LY, UK
patrick.w.hanks@gmail.com

Abstract. Word meaning is at best a very vague phenomenon – some lexicographers, including the present writer, have gone so far as to claim that word meanings do not exist. So how is it possible that people can achieve precision in the meaning of their utterances? And how is it possible to use language creatively, to talk about new concepts or to talk about old concepts in new ways? The answer is surprising; it calls into question most previous work in computational linguistics on the so-called 'word sense disambiguation problem', which, I shall argue, is still unresolved because it is based on unsound theoretical assumptions. If word senses do not exist, they surely cannot be disambiguated (or processed in any other way). The hypothesis to be explored in this paper is that meanings are associated with the phraseological patterns associated with each word in normal usage, rather than with words themselves.

Keywords: Lexicography · Corpus linguistics · Semantics · Word sense disambiguation · Phraseology

1 Introduction: The Need for Critical Examination of Received Theories

A guideline for researchers of any subject, whether grand or humble, could be this: "Descartes refused to accept the authority of previous philosophers." – Wikipedia.

In this paper I propose to explore some aspects of the phenomenon that is meaning in language. The hypothesis to be investigated is that meanings reside, not in words alone, but also in the phraseological patterns associated with each word. To put it another way, in natural language different aspects of a word's meaning potential are activated by different collocations. I shall illustrate this concept with corpus-based case studies of the verbs *grin* and *repair*.

Moreover, I shall argue that precision of meaning is achieved, not by a concatenation of words (in ways that have been traditionally envisioned for thousands of years by grammarians and logicians), but rather by a concatenation of phraseological patterns, the meaning of each of which is no more than a set of contrastive probabilities. At first, as pattern elements accumulate around a word in a text, its semantic entropy diminishes. But then, as a text becomes more extensive, entropy become resurgent. Considerable skill is required by writers of scientific and factual reports to keep entropy in perspective. The second law of thermodynamics says that entropy always increases with time. By

analogy, a law of dynamic semantics might state that semantic entropy always increases with text extent. The immediate environment of each word is critical.

As scientists and critics, it behaves us to be always prepared to re-examine the received theories of the past. This is particularly true of linguistics, for which a new kind of evidence suddenly became available during the last quarter of the 20th century, in the form of electronic corpora—very large collections of text in machine-readable form. This kind of evidence has supplemented and in many ways superseded previous ways of collecting evidence, which were of two kinds: evidence invented by introspection (researchers consulting their intuitions to imagine how they themselves use language) and citations painstakingly collected by hand. It is now clear that both these traditional kinds of evidence tend to introduce bias into the data. This is because both introspection and manual citation collection tend to favour the unusual, while failing to point out regular patterns of lexical co-occurrence.

Corpus evidence is not without dangers and biases of its own, but they are different in kind from the previous shortcomings, The main danger is of falling into the trap of believing that authenticity alone guarantees the existence of a meaningful pattern. It does not. In the words of Firth (1957), "We must separate from the mush of general goings-on those features of repeated events which appear to be part of a patterned process." Corpora enable us to do this.

From the 1980s onwards it became possible, for the first time ever, to use a computer to analyse statistically trends in patterns of usage. Statistical analysis of phraseology, in turn, opens up new possibilities for the analysis of meaning. It might have been the case that the new kind of evidence for phraseological norms would merely confirm received theories. Case studies show that this is not always the case. Occasionally, dictionaries get it wrong. For example, pre-corpus dictionaries typically define the transitive verb *file* as "to place (papers, records, etc.) in convenient order for storage or retrieval", while failing to notice the emergent role of this verb in talking about activating a procedure, as in phrases such as 'file a lawsuit', 'file a complaint', and 'file a flight plan'. These are very different meanings from putting papers into a filing cabinet.

Similarly, even the best linguists sometimes get it wrong. For example, Deignan (2005: 121) pointed out that the metaphor theorist Kövesces (1991) cites "Amusement gleamed in his eyes"—an invented example—as a realization of the conceptual metaphor HAPPINESS IS LIGHT. Corpus evidence, from both the Bank of English, which Deignan uses, and the British National Corpus (which I use) shows that Kövesces is right to imply that *eyes* is a normal collocate of *gleam*, but wrong to imply that it is normally associated with happiness. More regular collocates are *crazy, snarling, malicious,* and *mischievous*.

For reasons such as this, I would like to suggest that corpus evidence could and should have a profound effect on theories of meaning. Before going any further, therefore, I will attempt a brief summary of some salient points in the history of theories of meaning in European linguistics. Then we can discuss them.

2   Background: 100 Years of Linguistic Semantics in a Few Short Paragraphs

Throughout the 20th centuries, linguistics in the English-speaking world made steady and occasionally valuable contributions to the understanding of speech sounds (phonetics and phonology) and grammar (syntax and morphology). However, the study of meaning (semantics) was both confused and neglected. A distinction was made between lexical semantics and logical semantics. Lexical semantics was concerned with relationships such as synonymy, antonymy (various kinds of opposites), hyponymy (building hierarchies of semantic types on the basis of introspection) and meronymy (part-whole relations). It was not concerned with analysis of empirical evidence, such as the question of how people use words to make meanings. Logical semantics included topics such as Montague grammar, which I will not attempt to summarize here because, despite some hours of intensive effort since 1981, I do not understand it. Logical semantics also included truth-conditional semantics (very popular in the 1970s), which attempted to equate the meaning of a sentence with its truth conditions. The slogan of the truth-conditionalists was

> *"Snow is white" is true if and only if snow is white.*

This sort of thing is fine as a starting point for any philosophers of language who may be interested in distinguishing true from false statements of fact about the world. It is not useful for the study of language as a phenomenon of social interaction, and the attempt to convert the slogan into a statement about meaning (*"Snow is white" means snow is white,* sometimes unhelpfully extended to translations such as *"Snow is white" means la neige est blanc)* must now be seen for what it is, a resounding and content-free failure.

The European mainstream in linguistics goes back to Roman rhetoricians such as Quintilian (1st century CE) and grammarians such as Priscian (6th century CE). Quintilian was interested in meaning from a practical point of view – essentially, his job as a rhetorician was to teach people how to make effective and persuasive speeches. For this reason, his work is potentially of considerable interest to present-day semanticists and metaphor theorists, though few of them bother to mention him. Priscian, who lived five hundred years later, was the last and best of a long line of Latin grammarians, who saw it as part of their task to impose order on chaos, i.e. to ensure that the Latin usage of writers and public speakers was logical and as well-formed as a Roman sewer. For this reason, surviving Latin prose works, being written according to the rules developed by Priscian and his predecessors, are of great interest for the development of logic and grammar, but of very little interest to modern lexical semanticists. After Priscian, there is an eerie silence for 1200 years. Throughout the Middle Ages and the early years of the Enlightenment, European intellectuals were more concerned with issues in logic and theology than with natural-language semantics. Even grammar was neglected until the 19th century, when a great crowd of scientific philologists, including Wilhelm von Humboldt (1767–1835), Rasmus Rask (1787–1832), and Otto Jespersen (1860–1943), to name but three, burst upon the world, initiating the scientific study of languages and their history.

From the point of view of lexical semantics, the European mainstream runs from Humboldt through Ferdinand de Saussure (1857–1913) and the numerous German semantic field theorists of the 1920s and 30s. Humboldt's main importance lies in his focus on words and meanings and his recognition that all natural languages are rule-governed systems for making meanings—although since his time there have been (and still are, continuing to the present day) many furious disagreements about what sort of rules govern natural language systems. Gradually, it came to be recognized that Latin grammar is not the only kind of grammar that is possible and that other languages have different grammatical structures and even different word classes or components. (This message has still not reached some American and British dictionary publishing houses.) Saussure is widely hailed as the founder of modern scientific linguistics. His four basic dichotomies (langue vs. parole; and synchronic analysis vs. diachronic analysis; linguistic signs vs. what is signified; paradigmatic relations vs. syntagmatic relations) provide an essential foundation on which the analysis of linguistic meaning can be based. The relevance of these dichotomies to lexical analysis is discussed in Chapter 12 of Hanks (2013), which links Saussure's work to that of the semantic field theorists, mostly German, of the 1920s and 30s, and links them in turn to the Sapir-Whorf hypothesis, which argues that "the world is presented [to the individual human being] in a kaleido-scopic flux of impressions which has to be organized by our minds—and this means largely by the linguistic systems in our minds. We cut nature up, organize it into concepts, and ascribe significances as we do, largely because we are parties to an agree-ment to organize it in this way—an agreement that holds throughout our speech community and is codified in the patterns of our language." (Whorf 1940)

For purposes such as meaning analysis, the Saussurean theory of signs and its successors in various manifestations provided a foundation that needed to be comple-mented and integrated with a theory of valency. This was provided by the posthumously published work of Lucien Tesnière (1959), which underpins modern dependency grammar. The version of dependency grammar used in current on Corpus Pattern Anal-ysis (CPA) in English is colloquially known as SPOCA (an acronym for *Subject – Predicator – Object – Complement – Adverbial*), of which the best available account is Young (1980). The theoretical foundations of this grammar go back not only to Tesnière but also to Halliday (1961). Halliday's seminal article is the foundation of the systemic-functional tradition in linguistics, to which both CPA and distributional semantics are indebted.

To conclude this lightning survey of lexical semantics, one further name must be singled out, namely that of John Sinclair (1933–2007), a leading exponent, alongside Michael Halliday, of the empirical tradition in the European mainstream, which focuses on meaning as well as structure. Sinclair was the founder and editor-in-chief of the Cobuild series of publications in lexicography and English linguistics, adhering to a remarkably prescient plan that he set out in Sinclair (1966). His relevance to our present theme can best be summarized in the following quotation:

"Many if not most meanings require the presence of more than one word for their normal realization. … Patterns of co-selection among words, which are much stronger than any description has yet allowed for, have a direct connection with meaning." (Sinclair 1998).

Current corpus-driven work in English aims to give an empirically well-founded account of the patterns and to make explicit their connections with meaning. This will require, among other things, discovering reliable links between stereotypical patterns of word use and stereotypical meanings, of the sort described by Putnam (1975), which are very similar to the cognitive prototypes of Rosch (1975). This word (linking meanings to patterns of word use) is in its infancy.

3   The Purpose of Corpus Linguistics

What is the purpose of corpus linguistics? Why should we study and analyse empirical evidence for language use, rather than relying on our intuitions and knowledge of a language to provide examples?

To teach a class of students the difference between the active and passive voice, or singular and plural countable nouns in English, you don't need a corpus. Invented examples will suffice. However, corpus-based exercises can be useful for such activities as teaching the middle voice in ancient Greek - distinguishing the middle from passive. Does *luetai* mean 'he is set free' (passive) or does it mean 'he sets something free for his own benefit' (middle)? A corpus-based exercise could invite student to seek a collocating noun in the accusative case, which would mean that the verb must be in the middle voice. (I wish such an exercise had been available sixty years ago, when I was struggling with A Level Greek).

Corpus exercises may also be helpful for grammatical purposes such as reminding students of the many different ways in which German plural nouns are formed, or the distinction in different languages between countable nouns and mass nouns – for example to remind students that whereas English *money* is a mass noun (e.g. 'How much money have you got?'), Swedish *pengar* is a plural noun ('*Hur mycket pengar har du?*', literally 'How many coins have you got?').

Such pedagogical applications of corpus linguistics are useful enough, but they are trivial compared with 'the elephant in the room', namely the fact that corpus evidence constantly challenges some of our most basic received theories about language and meaning. To engage seriously with this problem, we need to analyse corpus evidence, i.e. the unbiased evidence of patterns of usage. A first step in this direction was the grammar patterns of verbs and nouns by Hunston, Francis, and Manning (1996, 1998).

4   Basic Patterns and Regular Alternations

It is now customary to divide the lexicon into 'function words' (determiners, auxiliary verbs, pronouns, prepositions, and conjunctions: so-called 'closed-class items') and content words (also known as 'open-class items': verbs, nouns, and adjectives). As with all linguistic categorizations, the boundary between these two classes is fuzzy, with some borderline cases. For example, prepositions are closed-class items that share certain properties—especially semantic properties —with open-class items, while the class of so-called 'adverbs' is a ragbag of heterogeneous lexical items, some of which have very

little in common. Content words (open-class items) are used to make meanings, while function words are used to show relationships among the content words.

A striking fact about content words (notably nouns and verbs) is that they are used in two quite different ways to make meanings. In the first place, they are used normally and conventionally, in syntactically regular valency structures with semantically regular collocates, whose regularity is immediately apparent in any large concordance to a corpus of ordinary texts. But in the second place, content words are occasionally used in unusual syntactic structures or with unusual, non-recurrent collocations. Such usages exploit the norms. Hanks (2013) demonstrates in some detail that exploitations are also rule-governed, but the rues for exploiting norms are not the rules of grammar. Instead, they are rules governing the creative use of ordinary language.

The remainder of this paper will focus on the corpus analysis of verbs, because the verb is the pivot of the clause. If we can get the phraseological analysis of verbs right, it is reasonable to expect that most of the nouns and noun phrases will fall into place. People interested in the phraseological analysis of nouns using corpus data are invited to consult Hanks (2000, 2004, and 2012), publications all of which have more to say about the corpus analysis of noun patterns.

One of the verbs selected for analysis, summarized in the *Pattern Dictionary of English Verbs* (PDEV), was *repair* (**see** Fig. 1). This provides material for our first case study, which will illustrate the relationship between subtly different patterns of normal usage, also the relationship between normal usage and exploitations. As is usual for PDEV entries, a sample of 250 corpus lines was extracted from the British National Corpus (BNC) and each line in the sample was analysed semantically and syntagmatically. The aim of such an analysis is to discover stereotypical patterns of usage and to map them onto stereotypical meanings. These stereotypical meanings take the form of so-called 'implicatures', which include entailments and presuppositions, among other things. A question that this procedure, called Corpus Pattern Analysis (CPA; see Hanks 2004, 2013; Hanks and Pustejovsky 2005) sets out to answer, for each sentence analysed, includes questions such as 'Who did what to whom?' (or, for intransitive verbs, 'What happened?'). A follow-up question, answered by the mapping of patterns onto implicatures, is, "How do we know?".

In the discussion that follows, the terms in double square brackets represent lexical sets that share a semantic type. They are listed in a hierarchical ontology of semantic types, publicly accessible as the CPA Ontology (http://pdev.org.uk/#onto). A lexical set may be vast, as in the case of [[Human]], which designates all the names of all the human beings who have ever lived, all the human beings who ever will live, and all the terms denoting a human being, including their different properties and roles, etc. Lexical sets consisting of only one word (or very few words) also exist; they are typical of idioms. An example of a lexical set of only one word is found in the idiom '[[Human | Institution]] grasp {nettle}'. The verb *grasp* is used in collocation with the noun *nettle* to make the meaning 'deal decisively with a difficult problem'. If someone talks about 'grasping the nettle', it is extremely unlikely (though possible) that they area talking about wild plants that sting passers-by (the usual meaning of the noun *nettle*). It is much more likely that they are using an idiom (which is nothing more than a conventionalized metaphor).

Fig. 1. PDEV entry for 'repair'.

Pattern 1 of *repair* is by far the most frequent, accounting for very nearly 60% of the corpus lines in the sample. Some basic stereotypes are: *repair a computer, an car, a building.* In these examples, the noun that is the direct object denotes a whole artefact, some part or parts of which are broken or damaged. These nouns alternate with others sets of nouns, each of which denotes a part rather that the whole artifact. These alternations include: *repair a keyboard, a wing mirror, a carburettor, the roof, the central heating system, a boiler.* In this set of examples we have the same pattern with the same meaning, but the focus is on the damaged part, not on the artifact as a whole. The implicature records a presupposition (that the [[Artifact]] or, more specifically, a part of it, has been damaged by some event) and an entailment (that the action of repairing puts the [[Artifact]] back into satisfactory working order).

Examples from BNC of terms denoting artifacts in the direct object slot, creating instances of pattern 1 at its most normal are 1, 2, and 3:

(1)   the owner's wish to *repair* the aircraft.
(2)   it took eight weeks to *repair* a motor vehicle.
(3)   a large company, with plenty of in-house techies to *repair* a computer if it goes wrong.

Examples from BNC of part-whole alternations in pattern 1 include:

(4)   *Repairing* the damaged cable took workmen about two hours.
(5)   Leaky taps are not expensive to *repair.*
(6)   Kalchu … began beating down the mud to *repair* the leaking roof.

Cable, taps, and *roof* are nouns that denote parts of a larger entity, rather than the entity itself. However, this difference does not justify making a separate pattern.

Pattern 2 ([[Human]] repair {damage}) is nothing more than a syntagmatic alternation of Pattern 1. That is to say, it has precisely the same meaning as Pattern 1. However, the focus is different, in that this pattern selects the word *damage* as the direct object, Thus, in this pattern, the presupposition has been selected as direct object, rather than the artifact that has been damaged. PDEV regards this admittedly subtle difference as sufficiently different to justify a separate pattern. In this way, the semantic integrity of lexical sets is preserved: despite regular occurrence in the direct object slot, we must recognize that damage is not an artifact; it is something that happens to an artifact. This must therefore be classed as a different pattern from Pattern 1.

Pattern 3 accounts for 8.8% of the 250 lines in our BNC sample. It denotes dealing with damage, not to a physical object, but to an abstract entity, namely a relationship among people or human institutions. This pattern relates to both Pattern 1 and Pattern 2. It could in principle be regarded as an exploitation of the first two patterns, in that a word that normally takes a physical object as its grammatical object is extended to an abstract entity. This sort of alternation (between physical object and abstract entity) is very common in English—and presumably other languages too. Classification into separate patterns is justified on grounds of frequency and regularity.

Pattern 4 (3.2% of the BNC sample) is a conventionalized metaphor. It is normal to talk of *treating* injured body parts, and unusual to talk of *repairing* a body part. By selecting the verb *repair*, the speaker or writer has assigned the human body the status of a machine. But with four or five examples in a 250-sentence sample, the metaphor can be regarded as a pattern (a norm or convention), rather than a creative exploitation.

Associated with pattern 4, we find two examples in the same paragraph of an unusual syntactic exploitation (7).

(7) His consultant was impressed with the speed that the arm was *repairing* in places. Mr Martinson was especially grateful for a modern medical process which may mean that his skin will *repair* without having skin grafts.

This is an intransitive use of what is normally a transitive verb. This may simply be a one-off exploitation. An alternative explanation is that it may be the tip of a domain-specific iceberg. Perhaps it is normal among doctors to talk of limbs and skin repairing, apparently without human intervention. The human body, after all, has a wonderful capacity for self-healing if left to itself in a favourable environment. These are questions that can be answered by assembling additional evidence for the use of this word, both general and domain-specific.

Finally, we come to pattern 5 (3.6% of our BNC sample). This is something completely different. It is a well-evidenced intransitive verb, whereas the preceding four patterns are transitive. And it requires an adverbial of direction governed by the preposition *to*. Pattern 5 is the only representation of the word *repair* in FrameNet (in the Self_motion frame). This means that FrameNet has nothing (yet) to say about 96% of uses of *repair*. This kind of problem is very common in FrameNet, due to its lexicographical methodology: it proceeds frame by frame on a speculative basis, rather than word by word in the light of corpus evidence. FrameNet never asks, "What are the different meanings of this or that word?".

To summarize: for the verb *repair,* PDEV recognizes four very closely related transitive patterns, plus an intransitive one that is very different.

As it happens, pattern 5 is classified by OED as a different word (a homograph), etymologically distinct from the senses of the other four patterns associated with this word. This distinction is made, not on the grounds of the very obvious syntagmatic or semantic differences, but on the basis of an extremely subtle difference in the spelling of the Anglo-Norman French source word(s): *reparer* in the case of the 'restore to good working order' senses, as opposed to *repeirer*, *reparier*, *repairir*, etc., for the verb of motion.

5   Different Kinds of Choice and Contrast

A contrasting verb is *replace.* Nowadays, for example, it is more normal to **replace** *a laptop than to attempt to* **repair** *it*. A language user makes a choice of the relevant word to express the intended meaning. *Repair* and *replace* denote choices of action the actant within the sentence (typically, the subject of the sentence). Such choices contrast with a different kind of choice, namely one that is made by the speaker or writer, not by a person involved in the event being described.

Repair, *mend, treat*, and *cure* express similar meanings: logically, one might say that the language does not need all of them. The choices are determined by context. Typically, you *repair* an artefact (a manufactured physical object that has a functional purpose), whereas you *treat* a human being or an animal. But *treat* does not imply that the treatment was successful, only that the attempt was made. For successful treatment, we have the verb *cure*. This is just one tiny example of the vast network of contrasts and choices offered by any natural language.

The contrasts just mentioned (*repair, replace, treat, cure*) constitute a paradigmatic set. Other contrasts resolve into meronymic sets, which are syntagmatic. For example, we may ask, what, typically, do you treat? The lexical set of nouns that includes the words *{patient, casualty, man, woman, child, animal}* as direct objects of the verb *treat* also includes words such as *{disease, illness, infection, ulcer, cancer, arthritis, diabetes, malaria}* and a vast number of other terms denoting diseases and ailments, which are meronyms denoting properties of patients receiving or requiring medical treatment. The Sketch Engine (https://www.sketchengine.co.uk/) is a tool that enables the user to instantly identify the statistically salient collocates of any word in almost any natural-language corpus. Associating these statistically salient collocates with particular meanings is a lexicographical task. In the current state of our knowledge, it cannot be automated. Attempts to do so have been premature. However, it is entirely possible that, given a large corpus of pre-tagged training data and advanced machine-learning techniques, automatic sorting of sentences into sense categories will before too long become possible.

6 Using Language Creatively

Creative use of language takes many forms. As far as the identification of patterns is concerned, it may involve unusual syntax, anomalous arguments, or figurative usage (Fig. 2).

Fig. 2. PDEV entry for 'grin'.

Our second case study illustrates the nature of creative usage. The verb grin is almost always intransitive. Apparently transitive uses, as in examples 8 and 9 (one with a cognate object, the other with a reflexive pronoun), may be regarded as syntactic exploitations of the intransitive norm. Because only one example of each of these syntactic structures was found in our BNC sample, they were easy to classify as exploitations of the norm.

(8) He *grinned* an uncomfortable grin.
(9) Ian McCann *grins* himself to death for a cosy tour round the home of Billy Bragg …

Other cases are more difficult to classify. CPA takes simple statistics seriously, so a group of frequent occurrences of a syntactic structure may be classified as a separate pattern in its own right if it is comparatively frequent, whereas, in the analysis of some other verb, the same grammatical alternation, if rare, may be classified as an exploitation of a norm. PDEV contains two alternations of *grin* as a verb that are separate patterns, plus one ("[[Human]] grin and bear it") that is an idiom with a particular meaning that has little to do with the core sense of *grin*. Thus, although PDEV records four patterns for this verb, there is only one basic literal sense, which is given first. Three of the patterns record subtly different semantics, each of which is associated with a subtly different phraseology.

The first of the regular alternation is the use of *grin* as a verb governing a speech act, as in 10.

(10)   'No,' she *grins*, 'this is far more sophisticated.'

The second regular alternation is the introduction of a second argument governed by the preposition *at*, as in 11:

(11)   We all shook hands and *grinned* stupidly *at* one another.

For convenience, PDEV treats both of these alternations as separate patterns, but they could in theory equally well be treated as optional arguments within the main pattern. This would, however, result in the patterns becoming unreadably complex, while giving undue prominence to minor variations. Any empirically valid analysis of lexis must focus on probable usage, not on all possible usage. Common sense demands that basic patterns should be simple, which may well reflect a cognitive reality.

Examples 12 and 13 may be seen as slightly unusual, in that the verb is complemented by an intensifying phrase ('from ear to ear' and 'like an idiot'), but PDEV does not consider these as separate patterns. These two sentences are classified as entirely regular examples of pattern 1 of the verb *grin*.

(12)   Kirkby comic Sean Styles is *grinning* from ear to ear with his latest job.
(13)   Henceforth I went around *grinning* like an idiot.

Finally, as we saw in the case of *repair*, natural language has mechanism for focusing on a particular aspect of what is said. One example will suffice. The most normal pattern for this verb is that *people grin*. A body part is not a person. However, it is quite unremarkable to select the relevant body part (the face or mouth) and focus on that as the subject of the sentence, as in 14.

(14)   The faces *grinning* from the tatty punk regalia looked … depraved and hollow-eyed.

There is a vast literature on metaphor, but the focus on metaphor is misleading in two ways. Firstly, metaphors are not the only kind of figurative language that exist. Chapter 8 of Hanks (2013) identifies several ways in which normal phraseological patterns can be exploited. These include not only metaphors, but also anomalous argument (not all of which are metaphors), similes, exaggeration (hyperbole), understatement (meiosis and litotes), euphemism, sarcasm, irony (saying the opposite of what you mean), and other tropes.

Secondly, as suggested in Bowdle and Gentner (2005), metaphors are of two kinds: conventional and creative. Conventional metaphors are nothing more than secondary conventions. The vast majority of metaphors are conventional, e.g. 15.

(15)   the *antibodies* that are the *front-line troops* of the immune system.

When I first read this, it struck me as a fine example of a freshly created metaphor. I was soon disillusioned. A few seconds with Google were enough to show that, in the literature of immunology.

7   Psycholinguistic Implications

Corpus pattern analysis is an exercise in sociolinguistics. The corpus lexicographer compares hundreds of different recorded uses, by many different writers and speakers, of the same word or phrase, and attempts to discern some underlying patterns of linguistic behaviour shared by many members of a speech community. Each of these observed patterns of linguistic behaviour is associated with a particular meaning: more precisely, with a particular set of implicatures, including presuppositions and entailments. On this basis, predictions can be made about what the word or phrase will mean when used in the same or similar contexts by anyone at any time in the future.

The corpus pattern analyst is not interested in trying to predict all possible meanings of a word or phrase; instead, he or she focuses mainly on predicting the *probable* meaning of the word or phrase when it is used in different contexts.

How does this work? According to Hoey's theory of lexical priming (2005), every member of a speech community is primed from the very first moment of exposure to linguistic utterances to recognize patterns of word use and to associate each one with a particular meaning or set of meanings. Moreover, he or she will also learn to make meanings, expressing his or her needs, observations, and questions by uttering different patterns and by ringing the changes on them.

In this way, every member of a speech community has got, stored in his or her brain, a subset of the normal patterns of word use in the language. When a person is listening to or reading a text uttered by someone else, some form of pattern matching is going on inside his or her head—matching the patterns uttered against the set of patterns stored, and seeking the best match. When a match is found, a hearer's meaning is activated. This may be something logical such as an entailment, but it can equally often be an emotion or evocation of an experience.

The patterns are recognized, among listeners and readers, for purposes of understanding at least some of what is said or written. The patterns are also available to the individual in order to express his or her needs, observations, and questions. Some people are very good at pattern recognition; others are better at pattern activation (without necessarily taking much notice what other people way or write). As a result, it is possible —indeed, perfectly normal—for users of the same language, who may never have met each other before, to achieve the mutual satisfaction of believing that they understand each other.

Societies have effective and sometimes brutal ways (including, principally, mockery) of eradicating anomalous uses of phraseological patterns and beliefs about their interpretation. Despite this, language change proceeds inexorably. New patterns and beliefs become established, old ones are driven into literary obsolescence, as each succeeding generation takes over the language for its own purposes and begins to elbow the older generation off the planet. Figurative language is a major component of the process of language change. Most figurative uses die as soon as they are born, but occasionally one will catch on and become established. Thus, today's metaphor may become tomorrow's literal meaning. It is in this way that the word ***dope*** progressed in the twentieth century from denoting a toxic-fume-emitting varnish used to stiffen the canvas fuselages of early aeroplanes (still used, I believe, on model aircraft) to denoting a

stupefying drug, and hence a stupid or stupefied person. An unusually rich set of other senses of this word are recorded, including "a thick gooey liquid used as a lubricant for skis and snowshoes or as food". The entry currently available in OED on-line is a mess, which is not surprising, as it was first published in 1897 and tinkered with, rather unsatisfactorily, in 1989. This was a period during which the word was undergoing sequences of ever more extraordinary, even baffling, meaning changes. It will be interesting to see what the OED lexicographers make of it when they get around to revising it, which cannot be far away.

8 Conclusion

Almost all content words (in all languages) are associated with a variety of different phraseological patterns, which are used by speakers and writers to create meanings. Patterns are established in a language over long periods of time, and (especially in a literary language such as English), they are fairly stable; i.e. a natural language contains a central core of phraseological patterns, which changes only gradually. Most of the phraseology used 200 years ago by writers such as Jane Austen is still in use today. Nevertheless, phraseological patterns do change; the changes are always unpredictable and sometimes sudden. An example that is quietly becoming established right now is the use of the preposition *of* in quantitive questions such as "How big of a bank loan do you need?" which would have been regarded as ungrammatical until the beginning of the present century.

The advent of corpus linguistics in the 1980s has provided us with a vast wealth of possible new ways of studying the extraordinary phenomenon that is natural language. But during the past twenty or thirty years, linguists have been remarkably slow to respond to the challenges. Largely for practical reasons, exacerbated by constraints on research funding, they have confined themselves to developing practical applications – in particular, practical pedagogical applications.

But now, the challenge is to take the opportunity to examine in detail the mechanisms of meaning in natural languages and to test any theoretical conclusions that emerge from such examinations against the evidence of actual usage, which can only be done with the aid of a computer. The theoretical foundations for such examinations were laid (on a necessarily speculative basis) in the second half of the 20th century by philosophers of language such as Wittgenstein and Putnam, literary critics such as I. A. Richards, and anthropologists such as Eleanor Rosch.

The dream of the Semantic Web (Berners-Lee, Hendler et al. 2001) is to "enable computers to manipulate data meaningfully." Up till the time of writing (September 2017), work on realizing the dream has done little more than construct ontologies and add tags to documents and elements of documents, to structure them and improve their machine-tractability, without engaging fully with their semantic content. As Wilks (2008) put it, the Semantic Web is "the apotheosis of annotation—but what are its semantics?" In the long term, if the dream is to be realized, something very like a pattern dictionary will be needed to explicate the relationship between word use and meaning for every content word in every language in the world. This is not such an impossible

dream as it may sound, for much work in distributional semantics, in many languages, is already under way. What is needed is a unifying theory of lexical semantics that will motivate and coordinate these efforts.

One essential future step—perhaps it will not be taken in full earnest until a decade or two from now—will be to examine, in the light of corpus evidence, the theoretical speculations about the nature of language and meaning by the great minds of the past, from Wilkins, Descartes, and Leibniz to Fillmore, Lakoff, and Chomsky. These great thinkers of the past did not have the benefit of corpus evidence or, if they did, they preferred not to use it. Fillmore's use of the corpus was satirized by the late John Sinclair as 'the fish-pond approach" (first dream up a theoretical point by introspection; then go fishing in a corpus for an example that fits the theory; if it does not fit the theory, throw it back and fish out another one).

The generative grammarian Noam Chomsky has been more forthright in his rejection of corpus linguistics, characterizing it as an accumulation of data without theoretical consequences. Chomsky has also said, "The smart way to keep people passive and obedient is to strictly limit the spectrum of acceptable opinion, but allow very lively debate within that spectrum—even encourage the more critical and dissident views. That gives people the sense that there's free thinking going on, while all the time the presuppositions of the system are being reinforced by the limits put on the range of the debate." Nowhere is this more true than of the aggressively constrained field of generative grammar. Chomskyans like to keep university departments of linguistics passive and obedient. The spectrum of acceptable opinion in such departments is limited to speculation about syntax and phonology, based on the evidence of introspection. Chomkyan linguistics, which has so little to say about lexis and meaning, might justly be characterized as "an impoverished and thoroughly inadequate conception of language", which, as it happens, is how Noam himself characterized Saussure's work.

References

Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001)

Bowdle, B., Gentner D.: The career of metaphor. Psychol. Rev. **112**(1) (2005)

Deignan, A.: Metaphor and Corpus Linguistics. John Benjamins Publishing, Amsterdam/ Philadelphia (2005)

Firth, J.R.: Personality and language in society. Sociological Review 62. Papers in Linguistics, pp. 1934–1951. Oxford University Press, London (1957)

Halliday, M.: Categories of the theory of grammar. Word **17**(3) (1961)

Hanks, P.: Do word meanings exist? Comput. Humanit. **34**, 205–215 (2000)

Hanks, P.: Corpus pattern analysis. In: Euralex Proceedings, Université de Bretagne-Sud, Lorient, France (2004)

Hanks, P.: How people use words to make meanings: semantic types meet valencies. In: Boulton, A., Thomas, J. (eds.) Input, Process and Product: Developments in Teaching and Language Corpora. Masaryk University Press, Brno, CZ (2012)

Hanks, P.: Lexical Analysis: Norms and Exploitations. MIT Press, Cambridge (2013)

Hanks, P., Pustejovsky, J.: A pattern dictionary for natural language processing. Revue française de linguistique appliquée **10**(2), 335–368 (2005)

Hoey, M.: Lexical Priming. Routledge, Abingdon (2005)

Hunston, S., Francis, G., Manning, E.: Collins Cobuild Grammar Patterns, vol. 1: Verbs; vol. 2: Nouns and adjectives, HarperCollins, London/Glasgow (1996; 1998)

Kilgarriff, A.: I don't believe in word senses. Comput. Humanit. **31**, 91–113 (1997)

Kilgarriff, A., Rychlý, P., Smrž, P., Tugwell, D.: The sketch engine. In: Williams, G., Vessier, S. (eds.) Proceedings of the 11th Euralex International Congress, Université de Bretagne sud, Lorient (2004)

Kövesces, Z.: Happiness: a definitional effort. Metaphor Symbolic Act. **6**(1) (1991)

Putnam, H.: The meaning of meaning and is semantics possible? In: Putnam, H.: Mind, Language, and Reality: Philosophical Papers. Cambridge University Press (1975)

Rosch, E.: Cognitive representations of semantic categories. J. Exp. Psychol. **104**(3) (1975)

Sinclair, J.: Beginning the study of lexis. In: Bazell, C.E., Catford, J.C., Halliday, M.A.K., Robins, R.H. (eds.) In Memory of J.R. Firth, Longman, London (1966)

Sinclair, J.: The lexical item. In: Weigand, E. (ed.) Contrastive Lexical Semantics. John Benjamins, Amsterdam/Philadelphia (1998)

Tesnière, L.: Eléments de Syntaxe Structurale. Klincksieck, Paris (1959)

Whorf, B.L.: Language, Thought, and Reality. MIT Press, Cambridge (1940)

Wilks, Y.: The Semantic Web: The apotheosis of annotation - but what are its semantics? Inst. Electr. Electron. Eng. (2008)

Young, D.J.: The Structure of English Clauses. Routledge, Abingdon (1980)