SELECTED EXTRACTS:

Page 2:
The present book suggests that many cognitive prototypes can be usefully mapped onto phraseological prototypes—phraseology in which the words collocate more frequently than by chance.

Pages 4-5:
In linguistic analysis, many things seem obvious, and only some of them are true. Determining the ‘normal’ uses of words turns out to be hard—indeed, impossible without very large bodies of evidence and a theory of prototypical norms, while corpora occasionally throw up bizarre utterances that are implausible but nonetheless authentic:

(1) Always vacuum your moose from the snout up, and brush your pheasant with freshly baked bread, torn not sliced. (Example cited by Judy Kegl (personal communication, c. 1986, on first looking into a corpus), from The Massachusetts Journal of Taxidermy, quoted in an article in a corpus of Associated Press newswire texts)

Example (1) is cited from memory, as I no longer have access to that early corpus. It hardly needs to be said that this example deviates from norms of usage in several ways—for example, the noun moose is not a canonical direct object of the verb vacuum. Occasional bizarre citations such as these have combined to deter lexicographers and linguists alike from investigating obvious
questions such as, “What sort of thing do you normally vacuum in English—or is the verb normally intransitive?”

1.4 Exploiting Normal Usage

Before going any further, I will give an extended example of how a single word is used normally and how norms are exploited. If we ask, “How is the English verb hazard normally used?” corpus evidence shows that the answer “People hazard guesses” accounts for over half of English sentences containing this verb. But as soon as a reader sees this answer, counterexamples may start springing to mind. Now, it is a principle of corpus analysis that invented examples and counterexamples should not be treated as primary data. So let us look and see what other uses of the verb hazard can be found in readily accessible actual data. Looking at half the British National Corpus (BNC50), here is a summary of what I found.

The part-of-speech tagger used by the Sketch Engine corpus search tool (Kilgarriff et al. 2004) reports 50 verb uses of hazard in this half corpus. Six of them turn out to be tagging errors: they are nouns or modifiers, not verbs. For example, the expressions hazard assessment, hazard control, and hazard studies are all modifier + noun, not verb + noun. These are mistakes—but they are not unreasonable and not unusual. An error rate of 12% in part-of-speech tagging is about par for the course. Of the remaining 44 uses of this verb in the sample, 23 take the word guess or guesses as a direct object. In other words, the expression to hazard a guess is clearly a prototypical norm of English usage, accounting for approximately 50% of all uses of the verb. But what about the other 21 uses of this verb? They include the following:

- (4) Although Phillips hazarded some theoretical conjectures concerning the relationship he had uncovered, it was left to others ... to provide the analytical underpinning for this astonishing statistical regularity.
- (5) Perhaps we can now hazard an attempt at defining ‘a good reader’.
- (6) He does not hazard any opinions on how costs depend on the complexity of the service!
- (7) In the spirit of such theorising, should we limit our visual world to one of rapidly succeeding shapes and colours from which we hazard the inference that a leaping dog is in the vicinity?
- (8) I would hazard that the ratio of real balances to total private sector net worth is less than 1% ...
- (9) “My uncle,” said Wendy, expanding further on her family, “was Provost of Dumfries; he had a rather odd name – ‘Chicken’.” “Not Hen Chicken?” I hazarded, as this humorous diminutive was part of my family mythology.
- (10) Although one may from time to time admire people who hazard their entire company on one major throw, it has to be a risk that for most of us would be, hopefully, both an unnecessary one and certainly an undesirable one.
- (11) Without a clear military objective, and a principle strong enough to hazard lives for, America cannot hope to lead by hassling.

Examples (4)–(9) can be classified as exploitations of the primary sense of the verb, whereas (10)–(11) are examples of a completely different norm for the same verb. Why should we say this? On what facts in the text is this assertion based?

In (4)–(9), the direct object has the same semantic type as the prototypical direct object, guess. Conjecture is a near synonym for guess. Both words denote speech acts that represent mental events in the mind of the speaker. An attempt at defining, like a guess or conjecture, is also a kind of speech act.
(although more likely to be a ‘written speech act’ than a spoken one). *Opinion* and *inference* are also kinds of speech acts representing mental events. A *that*-clause (as in (8)) represents a speech act or a proposition underlying a speech act, while direct speech (as in (9)) is an overt expression of a proposition. The words *I hazarded* are deliberately chosen in (9) (rather than the more neutral expression *I said*) in order to exploit the implication that what is said is a matter of guesswork rather than a simple factual statement. Thus, the prototypical norm influences the interpretation of the exploitation. It can be seen that in (4)–(9) the semantic values gradually move outward from the prototypical norm *hazard a guess*. Some speakers of American English assert that for them, direct speech with the verb *hazard*, as in (9), is unidiomatic. Be that as it may, even in British English this is a comparatively rare way of exploiting the norm, though all the more effective because of its rarity.

By contrast, in (10) and (11) the direct objects belong to a completely different semantic type and activate a different sense of the verb, one that is a close synonym of *risk*, in a sense discussed at length by Fillmore and Atkins (1992). Further examination of corpus evidence shows that people hazard not only their lives and their companies or business enterprises but also their wealth, their reputation, and other valued entities. Here, what is hazarded is an object of value, and the purpose of hazarding it is to try to obtain some benefit.

The two norms are not entirely independent. As a matter of historical fact, the notion of hazarding a speech act developed as an exploitation of the notion of hazarding or gambling money or an object of value in a game of dice. A person’s ability to say true things is, in social terms, an object of value. But this is irrelevant to the meaning and use of the word in modern English. Very often, a modern sense of a word develops as an exploitation of an earlier sense. Sometimes the two senses coexist peacefully for centuries. In other cases, the new sense elbows the older one out of existence.

### 1.4.1 Lexical Sets and Semantic Types

Most uses of many words are clear-cut cases and can be classified as normal, while a few other cases are clearly exploitations of a norm. But to complicate the issue still further, there are often a few cases that lie in a gray area between the category of norms and the category of exploitations. As we will see later on, the same is true of the dividing line between one sense or pattern of use and another.

The essential points to bear in mind are that a word may have more than one meaning, each meaning is associated with one or more patterns of use, any meaning or pattern may be exploited in different ways, and the dividing lines on all of these analytic planes are typically fuzzy. Up to a point, the analyst has a choice of how to apply the apparatus and where to draw the dividing line. Various levels of delicacy can be achieved. Thus, in the case of *hazard*, it is possible to draw the line very tightly, as was done above, defining the norm only in terms of the phrase *hazard a guess*, which is typical. Alternatively, it is possible to draw the line more loosely and say that the norm is ‘hazard [[Speech Act]]’—which would be to sacrifice semantic precision for the sake of greater inclusiveness. Various intermediate positions are also possible. What is not possible is to claim that, because there are borderline cases, no such distinction exists.

Few verbs are associated as strongly as *hazard* with just one noun as the typical direct object. However, almost all verbs are associated with different groups of nouns, and these different groups very often pick out different senses of the verb. The verb *fire*, for example, in one of its senses, is associated with
a set of lexical items denoting firearms: people fire guns, rifles, pistols, revolvers, machine guns, Kalashnikovs, and so on. A group of words like this is said to constitute a lexical set in relation to the verb fire. The lexical set is united by a common semantic type—namely, they are all firearms. A lexical set of this kind is given a name—the name of the unifying semantic type—which is conventionally written in double square brackets with initial capital letters, thus: [[Firearm]], [[Physical Object]]. Semantic theorists have tried to build semantic types into hierarchical structures called ontologies, discussed in Hanks and Jezek 2010. Thus, a [[Firearm]] is a [[Weapon]] is an [[Artifact]] is a [[Physical Object]], and so on. This lexical set and its semantic type contrast with other uses of the verb fire: for example, ‘[[Human]] fire [[Human]]’ (meaning ‘dismiss from employment’) and ‘[[Human]] fire up [[Machine]]’ (meaning ‘start’).

This may seem straightforward enough, but in fact the relationship between lexical sets and semantic types can be extremely complex. It will be a recurrent topic, from many different angles in many different contexts, in the chapters to follow. This book is not a theoretical account of hierarchical ontologies; it is an attempt to develop a different kind of theory, accounting for how words in use and meanings interact. The only thing that can be said with certainty about semantic types and ontologies is that, although at first they may seem like a matter of simple common sense, attempts to use them for tasks such as word-sense disambiguation have up to now yielded disappointing results. This is at least in part because the extent of fuzzy boundaries between categories of many different kinds has been severely underestimated.

### 1.4.2 An Extreme Example of an Exploitation

To close our first extended corpus-based analysis, let us return to the verb hazard and look at a more difficult example of an exploitation.

(12) I hazarded various Stuartesque destinations like Florida, Bali, Crete and Western Turkey.

To my surprise, I have found that many readers—especially computational linguists and other people with a logical orientation—coming to this sentence out of context like this judge it to be crazy, meaningless, unidiomatic, or uninterpretable. But this fact merely underlines the unnatural nature of what linguists and logicians do in general and what corpus linguists do in particular. No normal reader takes a sentence from the middle of a text and pores over it, without reference to what has gone before. Texts have a beginning, a middle, and an end. Example (12) comes from Julian Barnes’s 1991 novel Talking It Over. Barnes is a writer admired for his stylistic elegance—The Complete Review, for example, when reviewing this novel called him “a very fine stylist”—so our problems with interpreting this sentence are unlikely to be due to infelicity or ignorance of the language on the part of the writer. In fact, when the sentence is put back into context, it makes unremarkable good sense, in a way that can only be explained in terms of exploitations of norms. An extended context is given in (13):

(13) Stuart needlessly scraped a fetid plastic comb over his cranium. ‘Where are you going? You know, just in case I need to get in touch.’ ‘State secret. Even Gillie doesn’t know. Just told her to take light clothes.’ He was still smirking, so I presumed that some juvenile guessing game was required of me. I hazarded various Stuartesque destinations like Florida, Bali, Crete and Western Turkey, each of which was greeted by a smug nod of negativity. I essayed all the Disneylands of the world and a selection of tarmacked spice islands; I patronised him with Marbella, applauded him with Zanzibar, tried aiming straight with Santorini. I got nowhere.
At least two kinds of linguistic exploitation are present here. The first is ellipsis; the second is creative use of a combining form.

“I hazarded various Stuartesque destinations” is elliptical for “I hazarded a guess at various Stuartesque destinations.” Having just said that “some juvenile guessing game was required,” the speaker does not need to repeat the word guess. This particular exploitation rule (ellipsis) is then promptly repeated in five subsequent clauses, in each of which a noun denoting a location or type of location (Disneylands, spice islands, Marbella, Zanzibar, Santorini) is (in its particular context) elliptical for a speech act referring to a location. Moreover, a secondary exploitation of considerable complexity is involved in “tried aiming straight”: normally, you aim a gun straight at something, you aim (or fire) a question at someone; you don’t aim straight at a destination. However, it is noteworthy that, once the scenario has been set up, these stylistic complexities do not distract from the comprehensibility of the text. No ordinary human reader puzzles over what was being essayed or aimed at. This book shows why not.

The second directly relevant kind of exploitation that enables a reader to understand (12) is lexical creativity involving a combining form. Nowhere else in the novel, nowhere else in the BNC, and indeed nowhere in much bigger corpora—billions of words of English—have I been able to find the word Stuartesque. It is used occasionally in texts found on the Internet, with the meaning ‘characteristic of Stuart’, referring in each case to a completely different Stuart. But that does not make it meaningless or (in context) hard to understand. Stuart is the name of a character in the novel, and in English the combining form -esque is regularly affixed to a proper name to form an adjective meaning ‘resembling the person or place named, typically in respect of certain noticeable and even eccentric or bizarre characteristics’: Kafkaesque, Bergmanesque, Monroesque, Hollywoodesque, Dylanesque, Jaggeresque, Caravagesque are just a few of the examples of such creations attested in the Oxford English Corpus.

The verb hazard has been deliberately chosen for detailed expository discussion in the preceding paragraphs, as it is semantically comparatively simple, yet complex enough to allow illustration of the central argument of this book. Many words are more complex semantically—some much more complex—as we will see.
The reality is that Saussure gave us some thought-provoking though very sketchy observations about the probable nature of the lexicon, among other things; Chomsky has given us some thought-provoking though empirically ill-founded speculations about the possible nature of syntax and the human mind. Both sets of observations need to be examined rigorously in the context of a new science of empirical linguistics, and checked against corpus evidence, distinguishing normal usage from linguistic creativity. Only when this has been done will we be able to see whether integration is possible or indeed desirable. In this new science, empirical analysis of the lexicon will play a central role.

To be adequate, a linguistic theory must shed light on meaning as well as structure in language. In traditional lexicography, the picture as regards the task of explaining the conventions of meaning and use on which speakers and hearers mutually rely is blurred and distorted by at least the following factors:

- Inadequate criteria for linking meaning to use;
- Focus on rare and unusual words and meanings;
- Absence of statistically valid sampling;
- Failure to achieve appropriate levels of generalization, resulting (for example) in unnecessary and confusing duplication of senses;
- Undue influence of etymology.

Lexicography has long had a tradition of respect for evidence (in some but not all dictionaries), and learners’ dictionaries in particular now make substantial use of corpus evidence. However, lexicography has always tended to be theoretically weak, a situation that has not been helped by the palpable irrelevance to the lexicon of much linguistic theorizing. TNE [the theory of norms and exploitations], coupled with corpus evidence, provides the theoretical underpinnings that lexicography of the future will need. However, this will necessitate new kinds of dictionaries, capable of showing what each word means when it is used normally and what its normal uses consist of. In an ideal world, each account of a norm will be linked to a theoretical apparatus demonstrating how far the norms can be exploited creatively before meaning breaks down completely.

Effective lexical analysis depends crucially on interpretation of evidence through theory. Linguistic evidence is interpreted through a theoretical lens, but the theoretical lens must itself be ground and polished by data to ensure high precision; otherwise, our vision of the data will continue to be distorted, blurred, inadequate, and distracted by mirages.

Theories must be empirically well-founded. In the study of language, this means that examples supporting the theory must be selected from evidence of language that has been used authentically for some purpose, not on speculation and imaginative invention. Authenticity alone, however, is not enough. If a linguistic theory is to be of any use, it must not merely be supported by authentic examples; it must also provide a means of showing the regularities on which human linguistic interaction and communication depend, and of distinguishing these from deliberate irregularities.

To the extent that they do not meet this criterion, linguistic theories, including those discussed in this chapter, require validation in the light of evidence and in the light of a pervasive distinction between what is normal and what is creative, while always allowing for the gray area in between. No doubt, some theoretical insights, when tested against empirical evidence, will pass with flying colors. In such cases, nothing will need to be changed except perhaps the replacement of a few invented examples by real ones. At the other extreme, some popular linguistic theories will be found to be so remote from the realities of actual usage that they must be abandoned. The vast majority of cases
are somewhere in between. Important theoretical insights suddenly lurch into
error through the invention of evidence. Introspection is a necessary tool for
the interpretation of evidence, but it becomes worthless if it is used to invent
evidence.

No existing linguistic theory makes this distinction explicitly and systematically,
though some have hinted at it, and distinctions such as *langue* versus
*parole* and *competence* versus *performance* represent attempts to deal with the
obvious problems posed by obvious irregularities. But it is necessary to distinguish
deliberate irregularities from accidental errors. Performance errors in
speech, as speakers struggle to encode their meanings, are quite different from
deliberate exploitation of norms, which is itself a rule-governed procedure.

PAGES 410-408

CHAPTER 13: The Broader Picture

13.2 Summary of the Theory of Norms and Exploitations

. . . TNE is a ‘double-helix’ theory of language:
the set of rules that govern normal, conventional use of words is intertwined
with a second-order set of rules that govern the ways in which those norms
can be exploited and that contribute very largely to the phenomenon of language
change. As its name suggests, TNE is a theory with two main components.
However, unlike what we find with many other theories, these two
components are not sharply distinguished; rather, they are poles at opposite
ends of a cline. Some norms are more normal than others; some exploitations
are more outrageous than others. And in the middle are alternations: lexical
alternations, where one word can be substituted for another without change
of meaning (e.g., the idiom *grasping at straws* alternates with *clutching at
straws*); syntactic alternations (e.g., active vs. passive); and semantic-type
alternations, which are devices for selecting a different focus when talking
about what is basically the same event type (you can talk about calming
someone or alternatively, with a slightly different focus, about calming someone’s
anxiety; you can talk about repairing a car or you can focus on the
presupposition and talk about repairing the damage).

Above all, TNE is a theory of prototypes and preferences, based on extensive
analysis of actual traces of linguistic behavior—what people say and what
they may be supposed to mean—as recorded in large corpora. The lexical
analyst looks at large quantities of text data in various ways, using a variety
of corpus-analytic tools such as a KWIC index (a concordance) and statistical
analyses such as those provided by the Sketch Engine, and immediately perceives
that there are patterns in the way the words are used. Phraseological
preferences create phraseological prototypes, which map onto cognitive prototypes
of meaning and belief. More thorough analysis reveals further patterns,
hidden below the surface. The whole of a language is permeated with interconnecting
patterns. But as analysis of corpus data proceeds, something very alarming happens:
the patterns in a concordance that seemed so obvious and that caught the eye at first
glance begin to seem more and more difficult to formalize, as more and more unusual
cases are noticed. The difficulty lies in achieving just the right level of generalization.
More and more exceptions show up as the data accumulates, so in fact there is no single
right level of generalization, although it is only too easy to make generalizations that are
badly wrong.