A genre, collocational, and constructional analysis of RISK*

Donald E. Hardy and Crystal Broch Colombini
University of Nevada, Reno

Scholars of language have begun to explore the popular academic topic that is “risk” and to establish the viability of corpus linguistics for such analysis. We extend existing scholarship by analyzing the noun lemma RISK over the last two decades of contemporary American discourse with recourse to the Corpus of Contemporary American English. Using the corpus procedures of frequency analysis, concordance contextual analysis, collocational analysis, and a variation on distinctive-collexeme analysis, we revisit the often under-supported assertion that RISK is semantically negative, and then investigate the semantic potential for good risk, a phenomenon associated with economic and adventuring sectors. Data suggest that RISK is genre sensitive, predominantly negative, and heavily medical in shared meaning across genres, and further, that evidence for good risk is minimal. A phrase of interest in recent corpus work — RISK worth taking — is shown to be sensitive perhaps more to constructional constraints than to semantic prosody.

Keywords: corpus linguistics, risk, semantic prosody, lexical priming, distinctive-collexeme analysis, COCA

1. Introduction

In recent years, the cultural prevalence of risk, a heavily metaphorized and often contested concept, has increasingly intrigued practitioners from a multitude of academic fields. In addition to the disciplines and sub-disciplines focused specifically on the management, evaluation, assessment, prioritization, and communication of risk, many fields of study utilize and/or analyze the language of risk, although often in varied and even contradictory ways. Althaus’s (2005) extensive literature review, which uses the disciplinary mapping provided by the Encyclopaedia Britannica’s volume ProPaedia as a framework for a transdisciplinary analysis,
reveals not only the extent to which attention to risk has permeated the totality of academia — the only “glaring deficiency” (ibid.: 569) Althaus identifies is political science — but the broad range of epistemic stances on risk taken within it: “Each discipline applies a particular form of knowledge to uncertainty so as to order its randomness and convert it into a risk proposition” (ibid.: 569). As long ago as 1971, Ratcliffe attempted to set straight the field of insurance and risk on the meaning of risk, in part by specifying the precise meanings of “[u]ncertainty concerning loss” and “[p]robability of loss” (1971: 458–459). Ratcliffe’s paper demonstrates an approach that, although necessary in insurance and many of the fields where risk is managed, is prescriptive in attempting to regulate the understanding of risk. Our study is descriptive. We are interested in how corpora reflect the meaning, however “misused”, of the noun lemma RISK.

Disciplinary and professional use of the concept of risk has increased along several trajectories. In contrast to the scientific and mathematical approach to risk as calculable and objective are social scientific approaches that contextualize risk as a subjective phenomenon with substantial social and cultural implications. Sociology has been a particularly notable contributor to the conversation about risk’s social ramifications, with interest dating back more than three decades. In his 1984 Presidential Address to the American Sociological Association, Short’s underlying message was that risk was already at that time a hot topic — “An entire industry has grown up around risk analysis, complete with professional trappings, governmental agencies, and personnel drawn from a variety of disciplines and professions” (1984: 712) — and that sociological perspectives were badly needed to extend a too-narrow focus on human life, health, and economic values. Short’s call was answered. Diverse sociological perspectives on risk have since emerged, the most influential among them, arguably, the “risk society” thesis advanced by German sociologist Beck.

Beck identifies a break with modernity as society moves beyond classical industrial society to a new, reflexive modernity characterized as a “risk society”. Risk in the risk society is pervasive, unpredictable, and synonymous with danger. It is reflexive: society confronts itself through dealing with risk. Once calculable, it is now unpredictable. Once personal, it is now global as the consequences of industrialization are unloosed from their places of origin to diffuse widely, wreaking erratic consequences. The risk society replaces positive goals of wealth pursuit with negative, defensive goals of risk aversion: “Basically, one is no longer concerned with attaining something ‘good’, but rather with preventing the worst; self-limitation is the goal which emerges” (1992: 49).

Beck’s assertions regarding risk’s conceptual meanings have been broadly echoed in treatments of semantic meanings. In Risk, Lupton summarizes the scholarship on the several historical semantic shifts of “the word ‘risk’” (1999: 5–13),
arguing that at the advent of modernity, risk was “scientized” and treated in terms of calculable probability, as a means of evaluating knowable outcomes both desirable and undesirable: “‘risk’ could be both ‘good’ and ‘bad’” (1999:6–7). However, by the end of the twentieth century, the meaning was again shifting, with both the notion of calculability and most associations of the positive falling away. Lupton’s conclusions regarding contemporary uses of risk accord, in the main, with Beck’s: “In everyday lay people’s language, risk tends to be used to refer almost exclusively to a threat, hazard, danger or harm” (Lupton 1999:8).

Amidst a lack of detailed empirical evidence to support most generalizations about the contemporary meaning of risk, linguists have begun to undertake descriptive studies that attempt to account for definitional and semantic meanings of risk, often with the help of large corpora (Hamilton et al. 2007:164). Dictionary definitions would seem to support Lupton’s assertion of the term’s prevailing negativity. Fillmore & Atkins (1994) conducted a corpus-based frame semantics analysis of RISK as both noun and verb and explored the treatment of both the noun and the verb in ten dictionaries, as Hamilton et al. (2007:167) discuss in detail. For the noun, Fillmore and Atkins discover “at the most five dictionary senses”, all of which are primarily negative: “dangerousness, a dangerous situation […] unpleasant possibility […] someone or something dangerous […] the possibilities against which one can buy insurance […] the chance or hazard of commercial loss” (also qtd. in Hamilton et al. 2007:167). Only the last sense is potentially positive in that “there can be no success without risk”, as Fillmore & Atkins’s (1994:354) example entry reads. Fillmore and Atkins comment in a related corpus analysis that RISK belongs to the group of “English words whose semantic descriptions share some reference to the possibility of an unwelcome outcome” (1992:79; noted in Hamilton et al. 2007:166).

While recognizing the negativity of risk in general, some theorists and analysts point out that contemporary usage of risk is good at least some of the time. Lupton (1999:8) writes, “Risk is now generally used to relate only to negative or undesirable outcomes, not positive outcomes […]. In the esoteric parlance of economic speculation, however, there remain such things as ‘good risks’ in relation to making a profit”. Althaus’s review corroborates an equivocal perspective in economics, revealing risk as “a decisional phenomenon” comprising “a mix of challenge and security” (2005:573). Hamilton et al. (2007:178) indicate that “in the context of health and illness” RISK “has a negative semantic prosody even when that prosody seems neutral in other contexts”. However, they argue that the phrase risk worth taking “relates to positive connotations of taking risks and suggests a positive semantic prosody for the noun ‘risk’” (ibid.:176).

This paper assumes Hamilton et al’s (2007) perspective to investigate further the seemingly self-contradictory negative and positive semantics of risk. Our thesis is that uses of the noun lemma RISK are overwhelmingly negative, and that it
is important to examine the collocational and constructional details of both negative risk and so-called positive risk in order both to comprehend the ubiquity of negative uses of RISK and to understand the decidedly marginal existence of positive uses of RISK. In support of our thesis, we offer the following statistical and qualitative observations: (i) RISK can be understood as a genre-based phenomenon, concentrated in medically negative academic and magazine discourse; (ii) RISK generally is medically negative (see very similar conclusions in Hamilton et al. 2007); (iii) risk theorists posit good risk primarily in financial and adventuring sectors; (iv) good RISK is not a significant collocation; and finally, (v) phrases like RISK *worth taking* show preliminary evidence of being sensitive more to constructional constraints than to semantic prosody. Those constructional constraints themselves are motivated in part by the semantic features of negativity.

Our corpus-based study of the nominal uses of RISK builds upon and is heavily influenced by Hamilton et al.’s (2007) corpus study of all uses of RISK, both nominal and verbal. Hamilton, Adolphs, and Nerlich examine in detail the Collins WordbanksOnline English corpus (fifty-six million words) and the Cambridge and Nottingham Corpus of Discourse in English (CANCODE; five million words); they briefly consult the British National Corpus (100 million words) looking for examples provided by Lupton (1999). In their analysis of the Collins WordbanksOnline English corpus, they do not analyze specifically for the influence of genre on distribution, but the represented genres include British books, radio transcripts, newspapers, magazines, and transcribed speech, as well as American books and radio transcripts. They find both in random samples and in t-scores of collocations that RISK “often keeps the company of words from the context of human health and illness” (2007: 171). In their study, seven of the top forty collocates of RISK in the Collins WordbanksOnline English corpus are medical: cancer, disease, heart, health, breast, lives, and infection (ibid.: 168–171). All but one of these seven words (lives) are unambiguously nouns. In order to focus our discussion, we restrict most of our collocational analysis to nouns in this paper.

2. Corpus and methodology

Given the diverse meanings of RISK, data samples used for drawing inferences about contemporary meanings of RISK should arguably be large; given the assertion of evolving semantics, they should also be current. Thus, we began our investigation by attempting to corroborate the findings of Hamilton et al. (2007) using a corpus that is both significantly large and representative of published and broadcast contemporary American discourse — the Corpus of Contemporary American English (COCA), a relatively new body of texts (Davies 2008-). COCA’s
configuration lends itself to a genre analysis of very contemporary American discourse, from 1990 through 2010 (2010 only partially represented at time of final revision of calculations for paper, November 2010). COCA currently (November 2010) contains over 410 million words, each year of the corpus being represented by over twenty million words. Plans are to add twenty million words to COCA for each subsequent year (Davies 2009: 160). The sources in COCA are classified into five genres — spoken, popular magazines, fiction, newspapers, and academic journals — of approximately balanced representation (Davies 2009: 161–162). Within these categories, the great number of sources makes it difficult to assess any imbalance in word count among various publications within the specific domains. However, the corpus in general appears to be sufficiently representative of published and broadcast American English, with no indications that an over-abundance of data from any one source or type of source would skew data on general usage. Davies (2009:162) points out that “the texts within each of the five genres are balanced between the sub-genres”. Thus, “each year the newspapers are evenly divided between the ten newspapers [that Davies samples] (approximately 400,000 words each)” (ibid.).

We use four corpus procedures in this analysis: frequency analysis, concordance contextual analysis (in which we examine words in context qualitatively), collocation analysis (in which we examine the statistically significant co-occurring words), and an exploratory variation on ‘distinctive-collexeme analysis’. Sinclair (1991:170) defines collocation as “the occurrence of two or more words within a short space of each other in a text”. Gries & Stefanowitsch (2004:101) developed distinctive-collexeme analysis in order to explore “near-synonymous (or functionally near-equivalent) constructions” and “words appearing in particular slots in these constructions”. There are three commonly used statistical tests for strength of collocations, i.e. the measurement of the attraction between two words: log-likelihood, t-score, and mutual information (MI). Hamilton et al. (2007) use the t-score, while Davies’s COCA provides MI.

As Hamilton et al. (2007) acknowledge, both MI and t-score have limitations. MI overestimates the importance of collocations of low frequency, while t-score overestimates those of high frequency (Hamilton et al. ibid.: 171; for details on MI, see Manning & Schütze 2001: 178–183). For example, in COCA the collocation of primigravidas and RISK has an MI score of 8.37, which is quite high considering that the MI score for RISK and infection, a much more relevant collocate as we will demonstrate, is only 5.15. There are 470 examples of RISK and infection occurring within a five-word window of one another, while there is only one example of RISK and primigravidas occurring within five words of one another. Partially to compensate for the overestimation of the importance of low frequency collocations with MI score and because Davies’s Web interface for COCA provides MI
(not t-scores or log likelihood) for collocations, we consider for analysis only collocations with five or more occurrences. Manning & Schütze (2001: 182) indicate that a “cutoff” of three has been suggested to compensate for the overestimation of the importance of low frequency collocations in MI scores; they also say that “such a move does not solve the underlying problem, but only ameliorates its effects”. We have two reasons for maintaining a cutoff of five: first it “ameliorates” the overestimation of the importance of low frequency collocations, demonstrated above with the MI score for RISK and primigravidas; second, five is the figure at which we cease to see errors in the data surfacing in collocation counts. To illustrate, consider the sample in Table 1 from COCA of collocates of RISK with minimum frequencies set to one through four, in turn.

As is apparent, these collocates are simply errors in the data, errors that occur with such small frequency overall that they surface as “significant” collocates of RISK. While a cutoff of five does not guarantee the absence of data like those above, it does reduce the effect of such data on our analysis. Furthermore, only collocations with an MI score of 3.0 or higher will be considered of sufficient interest. The figure 3.0 has been determined to be a good cutoff value for clearly interesting collocates (Church & Hanks 1990: 24 set the value > 3.0; Davies 2008- sets the value ≥ 3.0 — cf. http://www.americancorpus.org/help/display_table_simple_e.asp?h=y).

### 3. Frequencies of the noun lemma RISK

Although there are clear indications that the frequency of the occurrence of RISK has increased over the last twenty to thirty years (Lupton 1999: 9–10), few studies apart from Hamilton, Adolphs, and Nerlich’s have paid attention to the influence of context (e.g. register or genre) in the distributional frequencies of RISK. In the second half of their paper, Hamilton et al. (2007) examine the five-million word CANCODE corpus, consisting mostly of informal spoken texts. They examine the

<table>
<thead>
<tr>
<th>Collocate</th>
<th>Frequency</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>treatedat</td>
<td>1</td>
<td>9.37</td>
</tr>
<tr>
<td>understood</td>
<td>1</td>
<td>9.37</td>
</tr>
<tr>
<td>age.this</td>
<td>2</td>
<td>10.37</td>
</tr>
<tr>
<td>substancegreat</td>
<td>3</td>
<td>10.96</td>
</tr>
<tr>
<td>usegreat</td>
<td>4</td>
<td>11.37</td>
</tr>
</tbody>
</table>
effect of register on the frequencies of RISK (both noun and verb) and find that RISK occurs in highest frequency in conversation between friends and in pedagogic contexts, but as they say themselves, “the overall differences in the percentage score may be too small to form the basis for any robust statements specifically regarding the frequency variability” among the five levels of register that they investigate in CANCODE, that is, conversation among living partners, friends, colleagues, and service participants; and pedagogic lectures and training (2007: 172–173). They find conversation about risk both among intimates and friends to be about “degree of risk”. In the pedagogic context, most examples of RISK occur in a medical context (e.g. medical lectures), a finding that tends to support the conclusion that RISK is mainly health oriented (ibid.: 176–177).

Even though Hamilton et al. find concentrations of medical collocates of RISK occurring in the medically pedagogic register in CANCODE, one might wonder about a more explicitly balanced collection of academic texts. And even though seven of the top forty collocates of RISK in the Collins WordbanksOnline English corpus are medical (Hamilton et al. 2007: 170–171), those collocates are for the aggregated corpus. That is, we do not know how widely shared among the genres particular collocates are. One of the primary reasons we find COCA useful for this study is that it affords the possibility of determining frequency and shared collocation in five balanced genres. Figure 1, which is a representation of the frequencies of the noun lemma RISK in the five genres of COCA, illustrates dramatic differences among most of these genres in the distribution of RISK.

The frequency of RISK in academic discourse is more than nine times higher than the frequency in fiction, while the frequency in magazines is more than

![Figure 1](image-url)
seven times higher. The frequency in magazines is also more than one and one-half times the frequency in either newspapers or spoken discourse, while the frequency in academic discourse is more than twice as high. Discourse on the noun lemma RISK, thus, appears to be heavily concentrated in academic journals and magazines.

4. Collocations of the noun lemma RISK

One of the best ways to determine the meaning of RISK in these genres is to investigate semantic preference, that is, “the restriction of regular co-occurrence to items which share a semantic feature” (Sinclair 2004: 32–33, 142). Although semantic preference can be supported with the occurrence of single instances of words indicative of a semantic category (ibid.: 32–33), we use collocational analysis in order to determine strength of preference. We do not concentrate our analysis of collocates in each of the genres individually. Instead, we attempt to determine to what degree the negatively medical or the generally medical is shared among the genres. If the medically negative is a core association of RISK, we should find evidence of the negatively medical across all genres in COCA. There are only five noun collocates within five words either side of RISK that occur with a frequency of at least five and that score at or above an MI of 3.0 in all five of the genres. Table 2 contains the frequencies and MI scores for these collocates.

Thus, if there are nouns that are highly typical of the semantic association of the noun lemma RISK, they are the nouns in Table 2. We will comment on the other words soon, but observe that the word *infection* hints at the inherent attraction of the noun lemma RISK not just for the medical but also for the negatively medical. To elaborate on our reliance on collocation to determine semantic preference, we note that Hoey (2005: 24) defines ‘semantic association’ (i.e. Sinclair’s ‘semantic preference’ 2004: 32) as the connection “in the mind of a language user” of a word “with a semantic set or class, some members of which are also collocates

<table>
<thead>
<tr>
<th>Collocate</th>
<th>Frequency</th>
<th>MI</th>
</tr>
</thead>
<tbody>
<tr>
<td>takers</td>
<td>62</td>
<td>5.81</td>
</tr>
<tr>
<td>factors</td>
<td>2,301</td>
<td>5.51</td>
</tr>
<tr>
<td>infection</td>
<td>470</td>
<td>5.15</td>
</tr>
<tr>
<td>exposure</td>
<td>418</td>
<td>4.20</td>
</tr>
<tr>
<td>risks</td>
<td>333</td>
<td>3.94</td>
</tr>
</tbody>
</table>
for that user”. By far the most common specific phrasing for the RISK and infection collocation is risk of infection, as in

(1) But there’s still the risk of a brain infection or the shrapnel that’s still in his head shifting [...] 
(COCA: Esquire 2006 (fiction); bolding throughout is ours)

Hoey posits ‘priming’ as a “property of the word” so that the mental activation of that word “may provoke” the mention of a cognitively related word (2005: 7–9). ¹ Priming itself as a property of the word (Hoey 2005: 8) or as a psycholinguistic “paradigm” (Anderson 1983: 87; Anderson invoked by Hoey 2005: 8) may be a result of neural “spreading activation”, roughly the activation of neural connections “from original sources to associated concepts” (Anderson 1983: 86–87; see elaboration and empirical testing of spreading activation, for example, in Roelofs 1992, Beer & Diehl 2001, and Erman 2007). Lexical priming is manifested here in the negatively medical by the word risk in risk of infection. On a textual level, lexical priming is highly evident when the same word or an inflected form of a word is repeated, frequently for cohesive purposes (Hoey 2005: 117). Another significant (as measured by MI) noun collocate of RISK is risks itself:

(2) Steere shrugged. “It was a risk, but I take risks every day.” 
(COCA: Cosmopolitan 1997 (fiction))

As we have shown, RISK occurs least commonly in fiction. Fiction has only sixteen noun collocates of the noun lemma RISK that score at or above an MI score of 3.0 with a frequency of at least five: taker, detection, infection, factors, risks, exposure, reward, risk, aids, counselor, gain, survival, danger, security, health, degree. By comparison, the spoken genre has 87, the magazine genre has 156, the newspaper genre has 76, and the academic genre has 183. Given the low frequency with which RISK occurs in fiction (cf. Figure 1), we exclude fiction from the data presented in Table 3, which contains the noun collocates within five words on either side of the

<table>
<thead>
<tr>
<th>General</th>
<th>Medical</th>
</tr>
</thead>
<tbody>
<tr>
<td>assessment</td>
<td>exposure</td>
</tr>
<tr>
<td>factor</td>
<td>downside</td>
</tr>
<tr>
<td>factors</td>
<td>risks</td>
</tr>
<tr>
<td>complications</td>
<td>taker</td>
</tr>
<tr>
<td>recurrence</td>
<td>takers</td>
</tr>
</tbody>
</table>
noun lemma RISK that occur in common among the remaining four genres and that score at or above an MI of 3.0 with a frequency of at least five.

The words in Table 3 are the core noun collocates of RISK in non-fictional discourse, which we have categorized as medical or general. These words give us a snapshot of the noun semantic associations of RISK in non-fiction. Those associations appear in their medical sense to be overwhelmingly negative, as in RISK of cancer/s, RISK of diabetes, RISK of disease, RISK of hepatitis, and RISK of obesity. Many of the collocates interact themselves; so for example in Table 3, words that might themselves seem potentially neutral, such as breast, lung, prostate, and colon have in their top ten collocates in each of the five genres individually the word cancer. Table 3 demonstrates that when medical collocates for RISK across genres are separated from more general collocates, or words that function in both medical and non-medical contexts (such as economics), over half of the collocates clearly belong in the medical semantic field. Consider the following two examples of the general collocate factors occurring in the near context of RISK. The first is from the context of business, the second from medicine:

3) Our Annual Report on Form 10-K, subsequent Quarterly Reports on Form 10-Q, recent Current Reports on Form 8-K, and other SEC filings discuss important risk factors that could contribute to such differences or otherwise affect our business, results of operations and financial condition.
(COCA: Associated Press 2007 (newspaper))

4) Researchers studying diabetes and heart disease already have picked the “low-hanging fruit” by identifying risks associated with such factors as age and gender, Marso said.
(COCA: Washington Post 2008 (newspaper))

Thus, while the language of risk certainly occurs in other discursive contexts, we have found strongly confirmed across all genres the trend noted by Hamilton et al. (2007) of a connection between risk and health.

Our frequency and collocation analyses have so far shown that the overall semantic association of RISK as a noun in non-fiction is negatively medical. Recall from Figure 1 that the noun lemma RISK appears to be sensitive to genre, occurring with highest frequencies in magazine and academic discourse. Table 4 demonstrates the even more overwhelming evidence for the negatively medical in the collocates of RISK in magazine and academic discourse. Table 4 contains noun collocates of the noun lemma RISK in those two genres that score at or above an MI of 3.0 with a frequency of at least five. Sixty-five percent of those collocates are unambiguously medical, and seventy percent of those medical collocates are clearly negative.
Our findings not only strongly support Hamilton et al.’s (2007) conclusions that RISK is associated with the medically negative but also demonstrate that both RISK and negative medical associations of RISK are unequally distributed across the genres of broadcast and print represented in COCA, with RISK and medical RISK concentrating in magazine and academic genres. The concentration of both RISK and negatively medical RISK in magazine and academic genres does not indicate that those genres determine the meaning of RISK. That concentration simply indicates that those genres have greater use for the negatively medical discourse of RISK.

Having now established the heavily negative semantic association of RISK, we next move to consider the question of whether or not RISK, in some contexts, shows evidence of positive meaning.

### 5. Research on good risk

In spite of the force of usage, as we describe it by means of COCA, and descriptions in dictionaries (most of those dictionaries, however, lacking the benefit of being corpus-based), arguments for the presence of good risk persist, coalescing around the entity of the risk-taker. Consider this passage, for example, from Lupton’s partial summary of the considerable discourse on the valorized risk-taker:

<table>
<thead>
<tr>
<th>General and Other</th>
<th>Medical</th>
</tr>
</thead>
<tbody>
<tr>
<td>risks</td>
<td>assessment</td>
</tr>
<tr>
<td>reduction</td>
<td>assessments</td>
</tr>
<tr>
<td>factor</td>
<td>complications</td>
</tr>
<tr>
<td>factors</td>
<td>recurrence</td>
</tr>
<tr>
<td>taker</td>
<td>aversion</td>
</tr>
<tr>
<td>takers</td>
<td>uncertainties</td>
</tr>
<tr>
<td>downside</td>
<td>occupant</td>
</tr>
<tr>
<td>hazards</td>
<td>interest-rate</td>
</tr>
<tr>
<td>exposure</td>
<td>radon</td>
</tr>
<tr>
<td>exposures</td>
<td>extinction</td>
</tr>
<tr>
<td></td>
<td>fracture</td>
</tr>
<tr>
<td></td>
<td>infarction</td>
</tr>
</tbody>
</table>
There is something extraordinary about the risk-taker. The person who attempts to sail or hot-air balloon around the world single-handed, the mountain climber scaling a dangerous peak, the daring entrepreneur who makes millions in business through taking risks, are all seen to stand out from the pack. (Lupton 1999: 155)

Our quantitative and qualitative investigation does not question the validity of Lupton & Tulloch’s (2002) and Tulloch & Lupton’s (2003) qualitative studies on voluntary risk-taking and the similar claims summarized in Lupton’s 1999 book (pp. 148–172), or even Lyng’s (1990) ethnographic claims about the attraction of “edgework”. In fact, as Tables 3 and 4 show, significant collocates of the noun lemma RISK in academic, magazine, newspaper, and spoken genres include both taker and takers. However, in order to highlight further the differences between what Lupton & Tulloch (2002: 115) term ‘dominant’ (negative risk discourse) and ‘counter discourses’ (positive risk), we turn now in our quantitative analysis of RISK to a search for what might seem to be an unambiguously positive use of RISK: good risks or good risk.

Whether Lupton means to use it in the literal sense or not, good risks is the phrase that she offers to illustrate positive risk, and she ties it to the particular genre of “economic speculation”, where high risk is associated with the potential for high return (1999: 8). However, good does not surface as a relevant collocate of either risk or risks in any of the genres of COCA, together or individually. In fact, in all five of the subcorpora the MI score for RISK as a noun lemma and good within five words either side of RISK is negative, meaning that the words tend to repel one another, although at relatively low levels: spoken (-2.43); academic (-1.58); magazine (-1.54); fiction (-.88); newspaper (-.80). The closer the score is to zero the more likely that the relationship between the words is due to sheer chance rather than attraction or repellency (McEnery & Wilson 1996: 71–72; Oakes 1998: 63–64, 89–90). Overall in the entire corpus as well, the MI score for RISK and good is negative (-1.75). In other words, the MI analysis shows that the words tend slightly to repel or tend to be “indifferent” to one another rather than attract one another. To place the unlikely combination of good and RISK into context, consider that the word good and the noun lemma PAIN with an MI score across genres of -1.03 repel one another a little less strongly than do good and RISK.

All of this is not to say that good RISK never occurs; in fact it does. The specific phrase good risk occurs twenty-seven times in the 1990–2010 COCA of 410+ million words, while the phrase good risks occurs four times. Of the twenty-seven occurrences of good risk, four are examples in which good modifies a complex noun phrase such as risk analysts or risk communication as in
(5) They had good risk analysts.
(COCA: Washington Post 2006 (newspaper))

Many other examples are hypothetical or negative as in

(6) If a business is a good risk, it will be able to get financing from the private sector; if not a good risk, it won’t get financing.
(COCA: Houston Chronicle 2000 (newspaper))

(7) He’s going to be 33 after the seasons [sic] over. It wasn’t a particularly good risk.
(COCA: USA Today 1997 (newspaper))

There are also financial examples that fit the usage Lupton posits:

(8) In one revealing if untrumpeted passage, the series noted that few bankers would rationally turn away a loan applicant they believe is a good risk.
(COCA: National Review 1993 (magazine))

Given that many appeals to risk in theoretical writings occur in the absence of either detailed statistics or even in some cases attested examples, it is important to examine the notion of positive risk more thoroughly. In the most obvious case of good RISK, as we have seen, there is a repelling or indifferent force between the words good and risk and good and risks. Further, we must point out that the MI score of bad and RISK is -.37, indicating that bad and RISK are pretty much randomly distributed with respect to one another (the MI score is close to zero). These behaviors parallel those of good and DEATH (-1.77) and bad and DEATH (-1.13). That is, in the genres represented in COCA, writers and speakers tend not to grade RISK or DEATH as “good” or “bad” for the reason that they are both semantically unmarked as bad.

6. Risk worth taking and positive semantics

Among Hamilton et al.’s (2007) findings, one of the most interesting to us was a sense of RISK that they argue “relates to positive connotations of taking risks and suggests a positive semantic prosody for the noun ‘risk’” (2007: 176). The authors report that this sense occurred with greater concentration in the transactional register (that which includes service participants). Examples given in their analysis are the following:

(9) <S 01> a risk worth taking Yeah it was a risk worth taking.

(10) of it but I thought it was a risk worth taking for er the opportunity
In (9), at least in the sample we are given, the item for which the risk is worth taking is not specified. In (10), that item is “the opportunity” for something that is not specified. A more explicit example of the same general construction occurs in COCA, as follows:

(11) School administrators feel the **risk** is **worth** taking for a once-proud program that has reached four Final Fours, but none since 1964.

(COCA: *Washington Post* 2006 (newspaper))

The gerund in **RISK (BE) worth taking** is similar to those gerunds identified by Huddleston (2002: 1246) to be licensed as a complement to the adjective **worth**. An example from Huddleston follows in (12):

(12) Your idea is certainly worth [giving some further thought to _____].

The adjective **worth** also allows an NP complement (Pullum & Huddleston 2002: 546), as we see in (13) and (14) below from COCA:

(13) But those who contributed to the National Academies report say preventing the suffering of people with mental illness is worth **the risk of some false positives**, partly because of the enormous cost of treating mental illness after it’s struck.

(COCA: *Time* 2009 (magazine))

(14) But at the end of the day, we believe that the risk was well worth **the possible reward**.

(COCA: *Fox Cavuto* 2004 (spoken))

We are especially interested in the remainder of this paper in constructions such as **RISK (BE) worth taking**, **RISK (BE) worth Y**, and **X (BE) worth RISK**, whether the complement to **worth** is a gerund (as in 9–12) or an NP (as in 13–14). Hamilton et al. (2007) argue briefly that the examples in (9) and (10), and others not given in their paper, are from speakers “who rationalize a particular course of action”. They explain that the examples are perhaps most understandable in the context of talk among strangers where “the lack of knowledge about the motivations and attitudes of the interlocutor may require a higher degree of justification of individual choices and actions” (ibid.: 176). Thus, if the examples in (9–11) reflect a positive semantic prosody for **RISK** they may be used in contexts, as Hamilton et al. argue, where the risk is felt to be positive by the speaker or writer and in need of justification.

We do not have space here to explore the history of semantic prosody and recent complex arguments about whether semantic prosody is a demonstrable linguistic phenomenon (Louw 1993; Sinclair 2004: 34–35; Whitsitt 2005; Hunston 2007; Bednarek 2008; see brief summary in Hoey 2005: 22–24). We will, however, take Louw (1993: 157), who is credited with the first published definition of
semantic prosody (Whitsitt 2005: 283), at his word in defining the concept as “[a] consistent aura of meaning with which a form is imbued by its collocates”. Thus, one could argue that at least part of the negativity of RISK is due to its frequent collocation with medically negative words such as infection, cancer, disease, and mortality. The positive semantic prosody for RISK that Hamilton et al. find in the RISK worth taking construction is at least not an obvious result of semantic prosody in the semantic preference (collocational) sense for the reason that there are no collocations to speak of within the construction other than RISK, worth, and taking.

De Smet (2009) makes a more syntactically- and discursively-based argument for the same positive semantics of RISK worth taking, appealing in part to construction grammar. Of such a construction as in (15), De Smet would argue that the subject of worth, which is risk, is also the moved object of the gerund taking, as in the underlying worth taking a risk:

(15) Given the scarcity of resources available to marginalized drug users in the inner city, repeated research participation comes to be seen as a risk worth taking.
(COCA: Journal of Drug Issues 2008 (academic))

De Smet argues that “[t]his redundancy [in the semantic merger of subject and object] allows valued and value to be integrated in one proposition, providing a single unified target for the speaker’s positive attitudes” (ibid.: 1738), in what he refers to as a ‘medial meaning’ analogous to that in middle voice conduives such as The car drives smoothly. De Smet argues that the “medial meaning” in constructions like risk worth taking, anything worth doing, everything worth knowing, and writers still worth reading (all attested in COCA) “no longer specifies the value of X as proportionate to the investment of time or energy required by doing Y, but signals that X is valuable and that in light of the value of X doing Y is also valuable” (2009: 1739). Thus, for De Smet, the “medial meaning” of RISK (BE) worth taking constructions would be inherently positive (e.g. his analysis of the risks were well worth taking 2009: 1737–1738). According to De Smet, a second discursive motivation for the X (BE) worth <gerund> construction is that subjects of worth that also have a gerund object are sometimes undervalued to the point that their real value is highlighted (2009: 1732, 1738), as in the following COCA example:

(16) Crispy onion strings, Caesar salad and baby back ribs are worth trying.
(COCA: San Francisco Chronicle 2008 (newspaper))

In this example, De Smet might contend, the food is perceived in a positive light because it takes so little effort simply to try the food. De Smet (2009: 1739) argues similarly that negative patterns sometimes reinforce by contrast positive patterns, as in it’s not a risk worth taking (COCA: E: The Environmental Magazine 2007
A genre, collocational, and constructional analysis of RISK

Although they are beyond the scope of our paper, De Smet’s medial *worth* meanings are deserving of further corpus research.

While we agree with De Smet and Hamilton, Adolphs, and Nerlich that the construction RISK (BE) *worth taking* is positive, we believe that constructional semantics are at least as responsible –if not more responsible than a syntactically-based “medial” semantic positivity or prosody– for whatever is positive about RISK (BE) *worth taking*. We will argue that the X (BE) *worth Y* constructions are constructional in Goldberg’s (1995; 2006) sense: that the meanings of the X (BE) *worth Y* constructions are not simply compositional sums of their parts (see Goldberg’s 1995:4 and 2006:5 definitions of ‘construction’). In particular, we believe that there is some evidence in COCA to support the claim that the subject and complement positions in the X (BE) *worth Y* constructions are subject to distinctive-collexeme analysis, an extension of collostructional analysis, as in Gries & Stefanowitsch (2004). Stefanowitsch & Gries (2003:214) explain, “[c]ollostructional analysis always starts with a particular construction and investigates which lexemes are strongly attracted or repelled by a particular slot in the construction (i.e. occur more frequently or less frequently than expected)”. Collostructional analyses performed by Stefanowitsch and Gries have included verbs in the ditransitive construction, the “result argument” of *cause*, and verbs “attracted” to the progressive construction (all and more in Stefanowitsch & Gries 2003). In their modification of collostructional analysis to include distinctive-collexeme analysis of verbs attracted variously to constructions that only appear to be alternatives of one another, they have investigated verbs attracted to the active vs. passive constructions, verbs attracted to the *will* vs. *be going to* futures, and modifiers attracted to the *gentive* s vs. *of* constructions (all and more in Gries & Stefanowitsch 2004). Gries and Stefanowitsch (2004:99) point out that patterns such as they examine in their 2004 distinctive-collexeme analysis have traditionally been viewed as alternational “partial” paraphrases. They argue that construction grammar is non-alternational, with distinctive-collexeme analysis in particular demonstrating the attraction of words with “semantic compatibility” for specific constructions. That is, in construction grammar, alternatives are not seen to be merely variations on basic syntactic patterns. As Gries & Stefanowitsch (2004:99) put it in summarizing the view within construction grammar as it is practiced by a number of linguists, Goldberg (1995; 2006) among them, “each member of an apparent ‘alternation’ is a construction in its own right […]; they appear to ‘alternate’ only because they are in a partial paraphrase relationship when used with certain lexical items”.

In order to provide a constructional analysis of RISK (BE) *worth taking* that is grounded in the context of other *worth RISK* and RISK *worth* constructions, we here extend our analysis to RISK (BE) *worth Y* and X (BE) *worth RISK*, which seem at least at first to be simple inverses, or alternatives:
(17) Would I make myself crazy before I made myself superhuman? The risk was **worth** the potential reward.

(COCA: *Bicycling* 2007 (magazine))

(18) If she attacked him, she would lose him. Venting her anger would shatter her future. It was definitely not **worth** the risk.

(COCA: *Webs of Power* 2008 (fiction))

Example (17) is superficially similar to the type in (9)–(10) identified by Hamilton et al. in that risk occurs before **worth**. However, it is unlike (9)–(10) in that it contains a noun complement to **worth**. Example (18) demonstrates the pattern for the overwhelming majority of sentences including **worth** as an adjective and RISK as a noun, where RISK occurs as a complement to **worth**.

Lawler’s (1989) explanation of **worth** in terms of a barter semantic frame will help us make initial explanatory sense of constructions such as those in (17) — that is, those in which RISK occurs before **worth** and which contain a noun complement to **worth**. In a “barter frame”, Lawler (1989: 391–392) argues, a participant wants one item more than another item; that participant must “have” one or the other of the items either before or after the barter (see also De Smet 2009: 1732). Lawler indicates that **worth** participates in the barter frame because the complement to **worth** states the degree to which the barter participant “wants” the subject in terms of a tradable commodity (the predicates HAVE and WANT are “primitives” in Lawler’s system (1989: 392, 386)). If “the car is worth the boat”, the barter participant wants the car at least as much or more than the boat. Lawler notes the degree of want for a subject with a $d$ and the degree of want for the complement to **worth** with a $d'$. He argues, in what we call the $X$ (BE) **worth** $Y$ frame, that $d$ is greater than or equal to $d'$, or $d \geq d'$ (Lawler 1989: 392). Thus, the unmarked relation $X$ (BE) **worth** $Y$ assigns greater value to the subject. In the $X$ (BE) **worth** RISK construction the benefit is equal to or greater than the potential of loss, represented by the risk, i.e. benefit $\geq$ potential of loss. So for instance, if we express the benefit in monetary terms (cf. Lawler 1989: 393) as either $100$ or $101$, and the potential of loss as $100$, the benefit is worth the risk in the sense that there is a potential of zero loss or a gain of $1$. Note that both $X$ (BE) **worth** RISK and RISK (BE) **worth** $Y$ inherently include chance in the equations as well as value, because as Fillmore & Atkins (1992: 81–82) point out, RISK involves both chance and value. The benefit is sometimes expressed as a given value, as in (18), where venting one’s anger is the benefit, and is sometimes expressed as a potential benefit, whose certainty is less than absolute, as in (17), where making oneself a superhero is literally the “potential reward”.

In the $X$ (BE) **worth** RISK frame, it is logical to refer to “wanting” $X$; however, it seems strange to speak of “wanting” risk. In RISK (BE) **worth** $Y$, one does not “want” the risk more than $Y$. These semantics are an indication that $X$ (BE) **worth**
Y might be a construction rather than a strictly analyzable assemblage of separate component parts; another indication is the status of *worth* as a “transitive adjective”, an “unusual” word category in English (De Smet 2009: 1732, n. 1). Pullum & Huddleston (2002: 546) list only four such adjectives in English, including *worth*. Returning to Lawler’s framework, we can see that if the subject and complement of the X (BE) *worth* Y construction are two objects of potential barter, then both X (BE) *worth* Y and Y (BE) *worth* X can be expressed as \( d \geq d' \). Thus if the car is worth the boat, then \( d \geq d' \), and if the boat is worth the car, then \( d \geq d' \) as well. However, note in (19) and (20) below what logically occurs in the equations for the X (BE) *worth* RISK and the RISK (BE) *worth* Y constructions:

\[
\begin{align*}
\geq & \\
19 & \text{The benefit is worth the risk.} \\
\leq & \\
20 & \text{The risk is worth the benefit.}
\end{align*}
\]

In (20), we see that the relational sign reverses precisely because RISK is *negative* in its unmarked sense. In RISK (BE) *worth* Y, the potential loss of the risk must be equal to or less than the gain of the benefit (Y). That is, (19) and (20) at least intuitionally appear to be simple alternatives, or partial paraphrases, of one another. We believe that our analysis below strongly suggests that they are not.

There remain at least two major questions in our study: first, why is the X (BE) *worth* RISK order (\( n = 363 \)) more than two and a half times as frequent as the RISK (BE) *worth* taking/Y order (\( n = 127 \))? A statistical analysis of a sample of nouns that appear in the X or Y slots supports the argument that those slots have different constructional meanings. We do not provide here a full distinctive-collexeme analysis (Gries & Stefanowitsch 2004) of the *worth* construction, in part because the construction is far more complex than it may look at first (for example with embedded nouns occurring in gerund complements to *worth*; e.g. *shooting elephants isn't worth running the risk* (COCA: *Newsweek* 1991 (magazine))) and in part because COCA is not syntactically parsed. Thus, in the limited space that we have here, it is impossible to provide an exhaustive operationalization of the enormous variety of syntactic patterns that occur in the subject (** worth**) and complement (** worth **) positions to *worth*. Instead of a strict distinctive-collexeme analysis, which would determine the level of attraction of every single noun, or noun lemma, that occurs in either the subject or complement positions, we provide a percentage ranking for attraction to either position of nine selected noun lemmas (in both singular and plural forms). Further, we are concerned here more with the relative negativity of RISK than with the details of *worth*, which are the subject of ongoing research by Hardy (in preparation). Consider the data in Table 5:
Table 5 suggests that noun lemmas with negative semantics are perhaps more attracted to the complement (worth ____ ) position and noun lemmas with positive semantics are perhaps more attracted to the subject (_____ worth ) position of worth. Thus, RISK occurs in seventy-four percent of its occurrences as the complement, along with high percentages of the other negative lemmas: EXPENSE, PAIN, and HASSLE. And LIFE occurs in eighty-three percent of its occurrences as the subject, along with high percentages of the other positive lemmas: BENEFIT, JOB, REWARD, FREEDOM, and PAYOFF. Recall that the MI score overestimates the importance of low-frequency collocations (Hamilton et al. 2007: 171, Manning & Schütze 2001: 178–183). Pedersen (1996) proves that the Fisher exact test as a measure of bigram strength not only does not overestimate the attraction of low frequency collocations but also does not rely on asymptotic assumptions of large samples (as explained in Stefanowitsch & Gries 2003: 218 and Gries & Stefanowitsch 2004: 101). Thus, the small frequencies of many lemmas occurring in constructions (collocations) are not an analytic problem for the Fisher exact test (as Stefanowitsch and Gries point out in their demonstration of the use of the Fisher exact test in collostructional analysis (2003) and distinctive-collexeme analysis (2004)). Our use of the Fisher exact test score for RISK is merely suggestive, for example, because it is by no means a complete distinctive-collexeme analysis (the data for our test include only the ten lemmas in Table 5). RISK, however, in these limited data is the most marked for its preferred position (as complement) with a probability score of 7.79E-67. With FREEDOM and PAYOFF removed from the calculations so that the “positive” and “negative” lemmas are balanced at four each, RISK remains most marked with a probability score of 2.86E-24.

Table 5. Example noun lemma distribution in _____ worth and worth _____ constructions in COCA

<table>
<thead>
<tr>
<th></th>
<th>_____ worth</th>
<th>worth ____</th>
</tr>
</thead>
<tbody>
<tr>
<td>RISK</td>
<td>127 (26%)</td>
<td>363 (74%)</td>
</tr>
<tr>
<td>EXPENSE</td>
<td>15 (20%)</td>
<td>60 (80%)</td>
</tr>
<tr>
<td>PAIN</td>
<td>12 (23%)</td>
<td>40 (77%)</td>
</tr>
<tr>
<td>HASSLE</td>
<td>11 (24%)</td>
<td>35 (76%)</td>
</tr>
<tr>
<td>LIFE</td>
<td>462 (83%)</td>
<td>96 (17%)</td>
</tr>
<tr>
<td>BENEFIT</td>
<td>73 (80%)</td>
<td>18 (20%)</td>
</tr>
<tr>
<td>JOB</td>
<td>76 (90%)</td>
<td>8 (10%)</td>
</tr>
<tr>
<td>REWARD</td>
<td>42 (80%)</td>
<td>10 (20%)</td>
</tr>
<tr>
<td>FREEDOM</td>
<td>19 (90%)</td>
<td>2 (10%)</td>
</tr>
<tr>
<td>PAYOFF</td>
<td>14 (78%)</td>
<td>4 (22%)</td>
</tr>
</tbody>
</table>
The final question for our study is what the discourse motivation is when a lemma that is inherently negative, like RISK, occurs in the “positive” subject (_____ worth) slot or a lemma that is inherently positive, like LIFE, occurs in the “negative” complement (worth _____) slot. We can do little more than hint at the many complexities here; Hardy (in preparation) contains further and complete details. However, we do suggest that the occurrence of RISK in the unexpected subject position is sometimes due to the newsworthiness of a relatively high valuation of RISK. This newsworthiness is signaled in part by the thematic importance of RISK as subject and therefore as the starting point for the predication (Halliday & Matthiessen 2004: 64), as in the following:

(21) The exercise is “an inoculation against racism,” she says. “We give our children shots to inoculate them against polio and smallpox, to protect them against the realities in the future. There are risks to those inoculations, too, but we determine that those risks are worth taking.”
(COCA: Smithsonian 2005 (magazine))

Previous to this passage, a controversial pedagogic exercise designed to combat racism has been described. Then, in (21), the exercise is metaphorized as an inoculation. Inoculations have their benefits, here evoked in the phrase to protect them against the realities in the future. However, note that the existent of the existential clause (There are risks) that begins the sentence with those risks are worth taking is risks (for existent and existential clause, see Halliday & Matthiessen 2004: 256–259). Those risks is then used as theme in the clause those risks are worth taking. One final small piece of evidence that RISK (BE) worth taking/Y is an expression of a newsworthy counter-expectation is that in three cases in COCA, RISK worth taking is a title or section head used within an article: e.g. “FOUR MONEY RISKS WORTH TAKING” (COCA: Men’s Health 2008 (magazine)).

In the opposite pattern, positive lemmas, such as JOB and LIFE, occur in the unexpected complement position (worth _____) to worth. In these cases, it appears that the complement position imposes a negative meaning of loss on these nouns because the complement to worth is always potentially a barter item (as in Lawler’s 1989 analysis). Stefanowitsch & Gries (2003: 213) similarly point out that “the ditransitive construction is said to coerce a transfer reading of hit” in constructions such as The coach hit the boy the ball, although hit has no inherent semantics of transfer. Thus, consider the following two examples:

(22) Is saving this two-foot-tall, 22-ounce, white spotted owl worth anybody’s job?
(COCA: CNN Crossfire 1990 (spoken))
7. Conclusions

Dingwall (1999), who critiques the risk society thesis and the failure of much social theory to ground its assertions about risk in empirical evidence, argues, “A core theme [of concern with negative risk and related negativity in critical theory] is the assertion that modernity has failed to deliver the freedom that it promised but has simply created new forms of enslavement, new obstacles to authentic communication” (ibid.: 476). The search for positive risk is then the inverse of the overestimation of negative risk. The high valuation of some risk seekers and the “risky” activities they engage in may indeed be a reality, but that group of privileged seekers of risk does not make risk good semantically. A detailed corpus analysis of the noun lemma RISK suggests the overwhelming linguistic background of the negativity of risk against which the cult of the daring risk seeker is narrated.

After confirming earlier arguments that RISK is primarily negative and heavily medical, our goal in this paper has been to examine what possibilities remain for the existence of good risk. We not only found RISK to be overwhelmingly negatively medical but found negatively medical RISK to be concentrated in magazine and academic discourse. Further, we found that “good risk” is exceedingly difficult to find, and that even when it looks as if we might have a positive semantic prosody for RISK in RISK (BE) worth taking/Y, it appears that RISK is still inherently negative although the construction _____ worth “coerces” (Stefanowitsch & Gries 2003: 213) some positive meaning from RISK. Further, it is actually a presumed negativity that is exploited in the newsworthiness of RISK (BE) worth taking. We found evidence in the constructions RISK (BE) worth taking/Y and X (BE) worth RISK that RISK patterns as an inherently negative word. We believe that it will be in the corpus linguistic context of examining constructions such as RISK (BE) worth taking/Y and X (BE) worth RISK that the inherent negativity and constructionally created positivity of RISK can be fully factored and understood.

In a recent issue of this journal, Barlow writes, “The lack of a clearly satisfying framework to account for collocational regularities is, I believe, one of the fac-
tors constraining the further development of a theoretical component to corpus linguistic analyses” (2011: 17). Barlow’s comments are made in the context of his discussion of the wide range of collocations and constructions that have been analyzed in corpus linguistics. Our analysis of RISK has focused on genre dependencies, collocations, and constructional possibilities. Future research in corpus linguistics and research on RISK in particular could profit from further exploration not only of the causes of regularities, perhaps Hoey’s lexical priming and Anderson’s spreading activation, but also of the causal interactions among (i) genre distributions; (ii) collocational patterns, such as RISK and the negatively medical; and (iii) preferred sequences, such as worth RISK (Table 5). That is, we have found distinct genre patterns and intermediate (collocational) regularities involving RISK, and there appear to be more immediate distinctive-collexeme regularities as well, all of which support a negative meaning of RISK. Remaining questions for future study include not only the strength and causes of interaction between genre, collocational, and constructional patterns of RISK, but also a determination of how widely spread those patterns are among other “positive” and “negative” lemmas.

Notes

* This paper has benefited from extensive discussion with Heather K. Hardy. We also thank three anonymous reviewers as well as one editor of IJCL for many very helpful suggestions. All errors of fact, theory, and analysis are our own.

1. Hoey (2005:178) points out later in his book that referring to “words being primed” is a “shorthand” for the complex interaction of individual and language: “[…] priming is what happens to the individual and is the direct result of a set of unique, personal, unrepeatable and humanly-charged experiences”. Furthermore, Hoey recognizes the interaction of context, individual language user, and lexis: “The contexts in which we encounter lexis contribute to the way it is primed for us and we are in turn and as a result primed to use such lexis in these contexts” (ibid.: 178).

References


Authors’ addresses

Donald E. Hardy
Department of English
University of Nevada, Reno
Reno, Nevada 89557-0098
donhardy@unr.edu

Crystal Broch Colombini
Department of English
University of Nevada, Reno
Reno, Nevada 89557-0098
ccolombini@unr.edu