7

Causal Attributions and Corpus Analysis

Justin Sytsma, Roland Bluhm, Pascale Willemsen and Kevin Reuter

1. Introduction

Most studies in experimental philosophy have used questionnaires involving vignettes. There are good reasons for the prevalence of questionnaire methods in experimental philosophy, including that these methods are fairly easy to use and are well-suited to investigating many of the philosophical questions that have been asked. As the present volume amply illustrates, however, questionnaire methods are not the only methods available to experimental philosophers, nor are they the only ones that experimental philosophers have used. In this chapter we will offer a brief introduction to a powerful set of non-questionnaire methods that can aid experimental philosophers in investigating a wide range of questions – methods of corpus linguistics.[1]

Our primary goal in this chapter is to introduce experimental philosophers to working with corpora, to survey some of the tools available and to demonstrate how these tools can complement the use of questionnaire-based methods. Toward this, we will put some of these tools to use in an area of research that has seen a flurry of interest in recent years – investigations of the effect of norms on ordinary causal attributions. Specifically, we focus on four questions:

Acknowledgements

We would like to thank an anonymous reviewer for the many insightful comments, as well as Eugen Fischer for his philosophical and administrative support.

[1] Although a handful of philosophers have used different tools of corpus linguistics (for a brief overview, cf. Bluhm 2016), and although the interest in the approach seems to have increased recently (e.g. Reuter 2011; Bluhm 2012; 2013; Hahn et al. 2017; Fischer and Engelhardt 2017; Sytsma and Reuter 2017 and the contribution by Mejia et al. in this volume – Ch.8), the use of corpora is still marginal to philosophy.

(a) Can corpus analysis provide independent support for the thesis that ordinary causal attributions are sensitive to normative information?
(b) Does the evidence coming from corpus analysis support the contention that outcome valence matters for ordinary causal attributions?
(c) Are ordinary causal attributions similar to responsibility attributions?
(d) Are causal attributions of philosophers different from causal attributions we find in corpora of more ordinary language?

We argue that the results of our analysis provide evidence for a positive answer to each of these questions.

Here is how we will proceed. In Section 2, we will briefly discuss recent work in experimental philosophy on ordinary causal attributions, highlighting our four questions. In Section 3, we introduce corpus linguistics. In Section 4, we bring corpus analysis methods to bear on our target questions. In Section 5, we use methods of distributional semantics to support the previous analyses. We conclude with some general methodological advice concerning the integration of corpus analysis techniques into experimental philosophy and philosophy as a whole.

2. Ordinary causal attributions and injunctive norms

Philosophical discussions of causation are often concerned with what has been termed *actual causation*. Actual causation is usually contrasted with *general causation*. A general causation statement describes a law-like relation between two types of events that stand in a causal relation, such as 'smoking causes cancer' or 'throwing rocks at windows causes them to break'. An actual causal statement, in contrast, describes the relation between two event tokens, such as 'Peter's smoking caused his lung cancer' or 'Jenny's throwing the rock caused the window to break'. For both general and actual causation, most philosophers assume that the concept of causation is a purely descriptive notion, referring to a relation in the world. As a consequence, a *causal attribution* such as 'A caused B' is true if and only if the relation of causation holds between A and B. Such an understanding of causation, however, means that normative considerations are irrelevant to causal attributions. The basic idea here is that whether or not an action is permitted by morality or convention simply does not matter for purposes of assessing whether that action, or the entity carrying it out, *caused* the outcome. Similarly, whether an action causes a morally good or bad outcome is irrelevant for causal considerations. Call this the *standard view* on causation.

Against the standard view, a growing body of empirical findings indicates that ordinary causal attributions are sensitive to normative information, prominently including *injunctive norms* (Hilton and Slugoski 1986; Alicke 1992; Knobe and Fraser 2008; Hitchcock and Knobe 2009; Sytsma et al. 2012; Reuter et al. 2014; Kominsky et al. 2015; Livengood et al. 2017a).[2] Injunctive norms include both *prescriptive norms*, which tell people what they should do, and *proscriptive norms*, telling people what they should not do.[3] Moral norms, such as the impermissibility of killing or hurting others, are prime examples of injunctive norms, but there is also a variety of non-moral norms that have similar action-guiding functions, such as social rules and regulations, etiquette norms and so on.[4]

Here are a couple of the empirical findings that have received much attention in the literature. Knobe and Fraser (2008) presented people with a story in which a secretary keeps her desk stocked with pens and both administrative assistants and faculty members help themselves from this stock. However, faculty members are not supposed to take pens. One day, both Professor Smith and the administrative assistant take a pen. Later that day, the secretary has no pen left to take an important message. Who caused the problem? In this case, Professor Smith and the administrative assistant performed symmetric actions (each took a pen), jointly leading to a bad outcome. The key difference between them is that while Professor Smith violated an injunctive norm (faculty members are not supposed to take pens), the administrative assistant did not (administrative assistants are allowed to take pens). Despite the two agents performing symmetric actions, participants were significantly more likely to agree that Professor Smith, the norm-violating agent, *caused* the problem than that the administrative assistant did.

To make matters more interesting, in a follow-up study, Sytsma et al. (2012) tested what happens if you remove the injunctive norm from the Pen Case, such

[2] These results do not directly challenge the standard view. Rather they put pressure on it insofar as philosophers are committed to what Livengood et al. (2017a) call the *'folk attribution desideratum'*. The folk attribution desideratum asserts that a key measure of the acceptability of an account of actual causation is that the verdicts it issues about specific cases line up with ordinary causal attributions about those cases. And there is reason to think that many, perhaps most, philosophers working on causation are committed to this desideratum.

[3] It should be noted that in the recent literature in experimental philosophy of causation, 'prescriptive norm' is often used indiscriminately to refer to both prescriptive norms and proscriptive norms as we understand them.

[4] Injunctive norms can be distinguished from descriptive norms (or 'statistical norms'). While there is an ongoing debate among experimentalists about whether descriptive norms have an independent effect on ordinary causal attributions (e.g. Knobe and Fraser 2008; Sytsma et al. 2012; Livengood et al. 2017a), we will focus on injunctive norms in this chapter.

that both Professor Smith and the administrative assistant are allowed to take pens. They found that participants now tended to *disagree* that Professor Smith caused the problem, while continuing to deny that the administrative assistant caused the problem.

Livengood et al. (2017a) used a computer case scenario, for which they found the same effects. More specifically, their studies revealed that participants were significantly more likely to agree that an agent who violated a norm caused a bad outcome, compared to the norm-conforming agent. Agreement that the norm-conforming agent caused the bad outcome was significantly below the neutral point, while agreement for the norm-violating agent was significantly above the neutral point.

While the Pen Case and the Computer Case are probably two of the most prominent examples in the literature, similar effects have repeatedly been found in subsequent research, and they seem to be robust for different causal structures (Kominsky et al. 2015; Sytsma et al. ms; Livengood and Sytsma under review), for both actions and omissions (Henne et al. 2015; Willemsen 2016; Willemsen and Reuter 2016) and across multiple test queries (Livengood et al. 2017a; Livengood and Sytsma under review).

Several explanations of the relevance of injunctive norms have been proposed in the literature. The most fundamental dispute regards the question of whether the observed effects reveal a real effect of injunctive norms on *causal attributions*. In two recent papers, Samland and Waldmann (2014, 2016) have denied this. According to their alternative explanation, when participants answer the question in studies like those noted above, they do not read the questions as being about causation, but about a related notion such as accountability or responsibility.

Those researchers who are convinced that norms do affect causal attributions have offered a variety of explanations of why and how this effect occurs. In the following, we will focus on two specific explanations, the *norm violation* and the *responsibility account*, as they make empirical predictions that we believe can be effectively tested with help of corpus analyses.[5]

The *norm violation account* put forward by Hitchcock and Knobe (2009) holds that the effects of norms in cases like we saw above are best explained in terms of the *cognitive processes* that lead to causal attributions being sensitive

[5] While we will focus on these two accounts here, it should be noted that these are not the only two plausible explanations in the literature nor are they the only two worth discussing. Just a few notable examples are the work of Alicke 1992, Cushman 2013, Malle et al. 2014, and Reuter et al. 2014.

to norms. According to Hitchcock and Knobe's account, causal judgements serve to identify suitable points for intervention in a system, and norms come into play because they affect which counterfactuals are most salient for determining the suitability of different intervention points. The basic idea is that in considering a situation, people think about how the outcome could have been prevented. But they do not consider every way in which the outcome might have been prevented; rather, they focus on those aspects of the situation in which something abnormal (i.e. counter-normative) has occurred. As such, while the norm violation account holds that the evaluation of norms is a crucial component in causal cognition, this does not mean that the concept of causation at play in ordinary causal attributions is a normative concept. Instead, norms come into play when people attempt to identify suitable intervention points in normatively laden situations.

The *responsibility account* put forward by Sytsma et al. (2012) contends that the concept of causation at play in ordinary causal attributions is itself normative. Thus, the cognitive process of making causal attributions starts off from a normatively laden concept. Instead of making a purely descriptive causal judgement that is later tainted by norms, the causal judgements are already normative. Sytsma and Livengood (2018, 7–8) express the idea this way:

> Saying that an agent caused an outcome […] typically serves to indicate something more than that the agent brought about that outcome or that the agent's action produced that outcome. Rather, it serves to express a normative evaluation that can be roughly captured by saying that the agent is responsible for that outcome or that the agent is accountable for that outcome, whether for good or for ill.

The norm violation account and the responsibility account make a number of diverging predictions. One such prediction is that while Hitchcock and Knobe hold that normative considerations have the same effect on causal attributions regardless of the outcome valence (regardless of whether the outcome is good or bad), the responsibility account allows that outcome valence might often make a difference.

As we noted above, Hitchcock and Knobe explain the influence of norms on causal attributions exclusively in terms of shaping the counterfactuals that are considered. And whether the outcome is good or bad would not seem to be directly relevant to assessing whether a counterfactual in which a candidate cause did not occur was more normal than what actually happened. Thus, Hitchcock and Knobe argue that to assess Alicke's (1992) competing account, which they read as explaining the influence of norms in terms of the desire to assign blame for the outcome, what is needed are cases in which a norm is violated and yet where no one is assigned blame because the outcome is good. Hitchcock and Knobe write that for such cases their account 'suggests that the impact of normative considerations *should remain unchanged* (because people still see that a norm has been violated)' (2009, 603; emphasis added).

The responsibility account, by contrast, does not make a direct prediction about the role of outcome valence in ordinary causal attributions; rather, it makes a prediction when coupled with a plausible prediction about responsibility attributions – that people are more likely to assert that an agent is responsible for a bad outcome than a good outcome.[6] If this is correct, and if causal attributions are relevantly akin to responsibility attributions, then we would expect that outcome valence will often make a difference.

One way to make progress on the issues concerning the role of norms on ordinary causal attributions that we have surveyed in this section would be to run still more questionnaire studies. However, we want to suggest that there is also benefit in approaching the problem from another angle. What we aim to demonstrate in this chapter is that another source of evidence can be brought to bear on these questions, namely corpus linguistics, and that its methods can both complement and enhance the use of questionnaire-based studies. After offering a brief introduction to corpus linguistics and applying some of its methods to the domain of ordinary causal attributions, we will return to the use of questionnaire studies and discuss potential shortcomings of such studies and how they can be alleviated through the use of corpus analysis.

3. The basics of corpus analysis

Corpus linguistics is a branch of linguistics that is defined by its use of *corpus analysis*. In its most basic sense, the term 'corpus' simply refers to 'a collection of texts' (Kilgarriff and Greffenstette 2003, 334) and 'analysis' to the process of

[6] We find this plausible because we expect that people are generally more concerned with assigning blame than praise, as a number of philosophers have noted. For instance, Prinz (2007, 79) writes: 'We blame someone for stealing, but we don't issue a medal when he refrains from stealing. We don't lavish the non-pedophile with praise for good conduct. In other words, we tend to *expect* people to behave morally.' While we will not argue for the veracity of this prediction here, it does find some support in the corpus analyses detailed below.

looking at the linguistic data that the corpus contains and assessing it for some research purpose.[7]

Briefly summarized, there are three sources of this method (*cf.* McCarthy and O'Keeffe 2010a). Its historical roots can be traced back as far as the Middle Ages, when concordances of Biblical words and phrases in context were compiled to assist exegesis. One of the basic functions of present-day corpora still is – particularly important for qualitative assessment of corpus data – to provide listings of queried linguistic expressions in context, in much the same mode of presentation as the one that mediaeval concordances used. The second important source of corpus analysis is the recognition of the importance of data representing actual use of a language, as opposed to data generated by the linguists themselves.[8] The third factor driving the development of corpus linguistics – of particular importance for quantitative analyses of corpus data – is the fast-paced development and spread of computer technology and the Internet in the late 20th century. Thus, while all corpora used in corpus linguistics are basically 'collections of texts', present-day corpora are typically collections of digitized texts that are accessed with computers.

Paradigmatic examples of well-known, large and freely accessible English-language corpora are the British National Corpus (BNC) and the Corpus of Contemporary American English (COCA).[9] The World Wide Web also offers a rich repository of digital texts, and while it is somewhat problematic to simply use the web *as* a corpus (termed WaC in the literature), there are some ways to tap into the Web's wealth of data by using extracts of it *for* building a corpus (termed WfC in the literature). We will make use both of COCA and a WfC approach below.[10]

One of the aspects that make a corpus out of a mere collection of texts is the decision to look within it for evidence of some use of particular linguistic expressions. The access to the data granted by the search engine is therefore not at all marginal to corpus analysis. Every run-of-the-mill search engine can do a full text search and handle wildcards, that is, it is possible to execute a query with a search string (a sequence of alphanumeric characters) in which some letter or

7 Helpful overviews of the discipline are given by Biber et al. 1998, McEnery and Wilson 2001, and McCarthy and O'Keeffe 2010b. For a quite comprehensive collection of articles on many aspects of corpus linguistics, see Lüdeling and Kytö 2008–2009.

8 See Leech 1992.

9 The number and variety of corpora compiled by linguists is ever-growing. Xiao 2008 and Lee 2010 give an overview of extant corpora, usefully sorted by type.

10 The useful terms 'WaC' and 'WfC' go back to de Schryver 2002. For a brief introduction to the rapidly growing field of 'web linguistics', you may turn to Bergh and Zanchetta 2008.

letters are substituted with a variable. For example, 'cause*' will not only find all instances of the use of 'cause' as a verb and a noun in the corpus, but also tokens of 'causes', 'caused' and more unexpected words such as 'causeway' and 'causer'. A more sophisticated corpus is pre-analysed and annotated with linguistic information, allowing, for example, to specifically search for the *lemma* 'cause' (all instances of the root word 'cause' regardless of inflection). It further allows to find tokens by grammatical category, for example, only instances of the verb 'causes' in its third person singular form, instead of the noun in its plural form; or co-occurrences of expressions, for example, the lemma 'cause' together with some noun (within a specified distance). Such queries are obviously much more powerful than mere full-text search. They are indispensable if a pertinent linguistic expression cannot be specified by a definite search string, or if, as in our case, the relevant linguistic phenomena include phrases such as 'responsible for' or 'caused' followed by some noun.

4. Exploring causal attributions with corpus analysis

We have already hinted at some very basic search options offered by common corpus search engines. Generally speaking, there are two approaches to using corpora (cf. Biber 2010): Corpus-*driven* research uses corpora to generate theories on linguistic phenomena from bottom up. Accordingly, the corpus is approached with minimal hypotheses as to the linguistic forms relevant to a given research question, for example, searching for tokens of 'cause' as a starting point to develop an understanding of causal attribution language. Corpus-*based* research, on the other hand, uses corpora to verify or falsify extant hypotheses about the use of language based on available theories about linguistic forms (for example, trying to confirm that 'cause' has some specific collocations). In practice, these approaches overlap to some extent, with researchers switching back and forth between them in their research process.

Similarly, corpora can be approached in a qualitative manner, that is, focusing on interpreting corpus findings with respect to meaning, or a quantitative manner, focusing on analyses based on countable objects and statistical facts. Both the methods used in this section and the next would count as quantitative on this definition. However, despite relying on frequency counts, the methods employed in this section have a somewhat more qualitative aspect to it, while the methods employed in the next are decidedly more quantitative, as will be apparent.

All the approaches to corpora we have hinted at depend on the existence of linguistic phenomena that can be traced with the help of available search engines. But in studying the language of causation, broadly construed, it is not immediately clear which linguistic expressions to look for. Linguists have identified an astonishing number and variety of ways that (arguably) are used to express causal relations.[11] Apart from the verb 'to cause' and the noun 'cause', as well as (partially) synonymous expressions, there are conjunctions for the subordination of clauses like 'because', 'since', 'as' (cf. Altenberg 1984; Diessel and Hetterle 2011), but also causative verbs, adverbs, adjectives and prepositions (cf. Khoo et al. 2002), and no exhaustive list of such means to express causal connection is available. There are also linguistic means to express a causal relation without lexical means, for example, the coordination of sentences and text organization (cf. Altenberg 1984; Achugar and Schleppegrell 2005). In consequence, it is only possible to find *some*, but not *all,* instances of causal language in a corpus with the help of a search engine. Moreover, most of the expressions mentioned above serve not only to express causal relations, but may also be used differently. To give but one example, 'cause' may also refer to a concern or purpose, as in 'her cause was just'.

With these caveats in mind, it is, of course, possible to access *some* of the causal language contained in a corpus. In our present context, we are interested in similarities and differences of the language of causal attributions and the language of responsibility attributions, and it seems plausible to approach our linguistic study with a focus on 'cause' as a relational verb and the phrase 'is responsible for', matching the type of phrases used in questionnaire studies to elicit causal attributions (e.g. 'Lauren caused the system to crash', 'Marcy is responsible for the death of the bystander').[12]

In the second section of this chapter, we surveyed some recent studies on ordinary causal attributions. These studies suggest that injunctive norms play a substantial role. Various theories have been proposed to account for this effect. We focused on two of these – the norm-violation account and the responsibility account. According to the norm-violation account, while ordinary causal attributions are influenced by norms, the underlying concept of causation

[11] We need not pass judgement at this point on whether such utterances really are intended to express or really do refer to a *causal* relation of some kind, let alone whether they express the specific relation at issue for causal attributions as we have defined them.

[12] The phrase 'is responsible for' is used to cue responsibility attributions in studies in Sytsma and Livengood (2018).

is descriptive and, thus, diverges notably from the concept of responsibility. Importantly for our research purposes, only norms, but not the valence and the severity of the outcome are said to affect causal attributions. From that we can infer the empirically testable prediction that the language of causal attributions and the language of responsibility attributions should be clearly distinguishable. In contrast, the responsibility account holds that the language of causal attributions and the language of responsibility attributions are quite similar. Moreover, the responsibility account predicts that the valence of the outcome will often have a notable effect on causal attributions.

If the ordinary concept of causation was a purely descriptive concept, then we would have no *a priori* reason to expect it to be used disproportionately in contexts with any particular valence. Rather, we would expect 'cause' to be used indiscriminately in the contexts where the outcome is good, in contexts where the outcome is bad, and contexts where the outcome is neutral. And if this was the case then we should see a good mix of positively and negatively connotated causal expressions without one of these types of expression dominating people's use of 'cause'. This is not what we find, however.

In order to investigate the nature of the terms used most commonly when expressing a causal statement, we looked at the most frequent nouns appearing after the causal phrase 'caused the' using the Corpus of Contemporary American English (COCA). The ten most frequently used nouns (numbers in brackets indicate the number of hits) are 'death' (103), 'accident' (87), 'crash' (80), 'problem' (79), 'explosion' (47), 'fire' (46), 'collapse' (27), 'injury' (26), 'damage' (24) and 'loss' (23). Independent judges classified all of these terms as negative.[13] Of the top 50 nouns, 30 were classified as negative, 19 neutral and only 1 positive. These data support the results from questionnaire studies indicating the relevance of norms to ordinary causal attributions. Furthermore, the commanding presence of negative terms strongly suggests that 'cause' is not only partly normative, but also primarily directed at negative outcomes. In other words, the results of

[13] Three independent raters were given a prompt – 'Please classify each of the following items based on whether you think instances of this type are most often positive, negative or neutral?' – followed by 780 items for classification. Items were the top 50 hits for 'caused the', the top 50 hits for 'responsible for the', and the top 20 hits for each of the eight synonymous expressions used below. This produced a list of 260 items that were then presented to each rater in three randomized orders. For the ten most frequently used nouns just reported, there was 99.4% agreement across the classifications with 169 out of 170 occurrences of these items being classified as negative. Overall, inter-rater agreement was high with a Kendall's tau of 0.751 and Spearman's rho of 0.803 averaging across the values for each possible pair of raters. For subsequent classifications we treated a term as negative (positive) if it was classified as negative (positive) at least two-thirds of the time.

our analysis strongly indicate that outcome valence has a substantial effect on ordinary causal attributions.

A key objection to our interpretation of these corpus data might be raised at this point. First, one important source of the corpus we have used is newspaper articles. And newspaper articles are notorious for focusing on negative events. Accordingly, it would not be surprising to find statements about causal relations for which the outcome is often negative. However, when we limited our search to sources of other types, like fiction, as is possible in COCA, no differences were found. For example, the top ten list of nouns following the phrase 'caused the' for spoken language only were 'death' (41), 'accident' (40), 'crash' (38), 'fire' (30), 'problem' (30), 'explosion' (26), 'plane' (11), 'damage' (9), 'recession' (9) and 'collapse' (8).[14] And a similar list was obtained when the corpus was restricted to academic texts.

It might be offered in rejoinder that a focus on the negative is simply a part of the human condition. As such, it might be suggested that terms that are arguably synonymous with 'cause' will also tend to be used in negative contexts. If that were true, then we could not conclude that we have discovered a specific characteristic of the language of causal attributions, but rather a more general phenomenon, for which an entirely different explanation would seem to be required. In order to investigate this objection, we posited the following null hypothesis:

> *Synonymy Effect*: There is no significant difference in the normative use between 'cause' and synonymous expressions.

If 'cause' is indeed specific in being used in a predominantly normative way, then we should be able to falsify Synonymy Effect. To test the hypothesis, we executed a corpus search with the eight terms that are listed by the English Thesaurus as being often used synonymously with the verb 'to cause': 'create', 'generate', 'induce', 'lead to', 'make', 'precipitate', 'produce' and 'provoke'. We inserted the phrase '*Φed* the' and noted the 20 most frequent nouns that appear after that phrase for all eight synonymous terms. (Table 7.1 lists those terms that were rated negatively for 'caused the' as well as the synonymous phrases.)

Considering only the 20 most frequent nouns, we calculated whether there was any significant difference in the use of 'caused the' compared to synonymous

[14] Only 'plane' was classified as neutral, with 160 out of 170 occurrences of these 10 terms being classified as negative by our raters.

Table 7.1 Most frequent negatively connotated nouns after phrases synonymous to 'caused the'.

Phrase	Number of negative terms (out of 20)	Negative terms
caused the	16	death, accident, crash, problem, collapse, injury, damage, loss, crisis, destruction, decline, extinction, harm, demise, explosion, fire
created the	3	problem, need, illusion
generated the	3	waste, war, killing
induced the	4	coma, panic, defendant, opposition
led to the	7	death, arrest, collapse, demise, loss, firing, end
made the	2	mistake, cut
precipitated the	13	crisis, war, attack, decline, conflict, collapse, downfall, fight, invasion, violence, demise, end, split
produced the	1	plutonium
provoked the	9	anger, fight, violence, murder, resignation, rebellion, strike, crisis, evacuation

expressions. Pearson's chi-square tests revealed that only 'precipitated the' was not significantly different ($\chi^2(1.13, 1) = 0.288$). All other comparisons were highly significant: 'provoked the' ($\chi^2(5.23, 1) = 0.022$), 'led to the' ($\chi^2(8.29, 1) = 0.004$); $p < 0.001$ for all other phrases.

The results are noteworthy in a couple of respects. (a) The searches demonstrate that the COCA corpus is not – at least not strongly – tilted towards texts that feature negative events. (b) The Synonymy Effect is likely to be false. Seven out of eight synonymous expressions of the form '*Φed* the' are not only significantly different in their most frequent uses compared to 'caused the', most of the synonymous terms seem to be used mainly in a neutral fashion. This means that the effect we recorded for 'cause' seems to be rather specific.

We have seen that Sytsma and Livengood propose that causal attributions are inherently normative, being used to assign responsibility. If this is correct, then we would expect responsibility attributions to be similar to causal attributions, including that they should also tend to occur more often in negative contexts. To test this expectation, we ran the same corpus search as before, this time entering the phrase 'responsible for the' and recorded the most frequent nouns that occur after that phrase. The ten most frequent nouns are 'death' (130), 'attack' (46), 'murder' (44), 'actions' (43), 'safety' (42), 'development' (42), 'loss' (35), 'design' (34), 'decline' (31) and 'violence' (31).[15] Of the 50 most frequent nouns occurring after 'responsible for the', 19 terms were rated negative, 17 neutral and 14 positive. It should be further noted, however, that many of the positive terms seem to belong to an alternative sense of 'responsible' from the responsibility attributions we are after – a sense indicating the duties involved in a role (e.g. 'content', 'creation', 'design', 'implementation', 'safety', 'security').

Nonetheless, the results show that 'responsible for the' has a similar environment to 'caused the' in being normatively laden and more often directed at negative events than positive. And by looking in greater detail at the respective numbers of hits for various terms, we found more striking similarities. For many terms like 'death', 'decline', 'damage', 'destruction', 'crisis', the use of responsibility language is roughly as frequent as causal language, suggesting that at least for some terms, both phrases might be used interchangeably (Table 7.2 lists the number of hits for these terms as well as the frequency ratio). This provides further support for the hypothesis that the causal attributions and responsibility attributions are often used to express the same state of affairs.

However, other comparative results between 'responsible for the' and 'caused the' do not quite match, which might suggest that we cherry-picked the data that fits our hypothesis. Table 7.2 lists two terms ('attack', 'murder') which are far more commonly used with responsibility language than causal language; for example, people seem to be far more likely to say 'she is responsible for the attack' than 'she caused the attack'. An explanation is easy to give: when we want to express a causal relation between a person and an attack or a murder, we would usually just rely on the causative aspect of the verbs and say that 's/he attacked' or 's/he murdered'. In other words, the English language has a simpler means to express causal language when it comes to attack and murder. The opposite result was found for the terms 'problem' and 'accident'. The corpus analysis revealed that 'caused the problem' is far more frequent than 'responsible for the problem'. If both concepts are akin, should we not expect that their uses are similarly frequent? Here, a closer look at the search hits is helpful.

What we find is that speakers often raise questions like 'what caused the problem?' leaving it open that it was not an agent but rather an event that

[15] Every occurrence of six of these terms was classified as negative by our raters. Every occurrence of 'safety' and 'design' was classified as positive. Classifications for 'actions' and 'development' were split, although they were positive overall. In total, 154 out of 198 occurrences of these 10 terms were classified as negative.

Table 7.2 Hits for some of the most frequent nouns after the phrases 'responsible for the' and 'caused the' and the ratio between them.

Word	responsible for the	caused the	ratio
death	130	103	1.26:1
decline	31	15	2.07:1
damage	24	24	1:1
destruction	18	16	1.13:1
crisis	11	17	0.65:1
attack	46	3	15.33:1
murder	44	2	21:1
problem	13	79	0.16:1
accident	11	87	0.13:1

caused the problem. In contrast, responsibility language is mostly used in relation to agents.[16] Thus, it is relatively rare that people make claims such as 'the malfunctioning brakes are responsible for the accident' but rather speak of malfunctioning brakes causing accidents. This in turn suggests that the semantic similarity between the language of causal attributions and responsibility attributions might be most pronounced for agent causation. A further investigation into this possibility is, however, beyond the scope of this paper.

It might be objected that the similar frequency in use of 'responsible for the' and 'caused the' for many nouns are merely coincidental and do not reveal any semantic similarity between these phrases; other phrases may be just as frequent. To counter this objection, we further examined which verbal phrases occur most frequently before nouns such as 'death', 'decline' and 'destruction', for which we have observed the same frequencies. The word 'death' was most frequently preceded by the verbal phrases 'caused the' and 'responsible for the'.[17] For the term 'decline', only 'contributed to' was a more common phrase than 'caused the' and 'responsible for the'. And for the term 'destruction' only 'stopped the' and 'prevented the' were more common than 'caused the' and 'responsible for the'. These data suggest that the similarity in use between the language of

[16] There is some quite strong evidence coming from further corpus analyses that support such a view. When entering the phrase 'what caused', COCA delivers 1,189 search hits compared to only 250 hits for 'who caused'. The situation is reversed for responsibility language. A search on COCA lists 412 hits for 'who is responsible for' but only 16 hits for 'what is responsible for'.

[17] In fact, 'seek the', 'face the' and 'get the' were even more common, but occurred not together with 'death', but mostly with the fixed expression 'death penalty'.

responsibility attributions and the language of causal attributions is unlikely to be a mere matter of coincidence. Rather, corpus analysis indicates that these languages are highly similar in meaning. This also puts pressure on Samland and Waldmann's (2014, 2016) alternative account of the effect of normative information on ordinary causal attributions: it does not seem to be the case that participants read questions in vignette studies to be about a related notion such as responsibility. Instead, the data suggests that the notion of causation is in itself inherently normative.

5. Corroborating the findings with distributional semantics

In addition to the somewhat qualitative approach to corpus analysis taken in the previous section, there is also a more mathematical way of exploiting corpus data by using an array of computational methods for investigating word meaning. One prominent approach is based on the 'distributional hypothesis', which follows Firth's dictum that 'you shall know a word by the company it keeps' (Firth 1957, p. 11; see also, Harris 1954). Accepting this, word meaning can be explored by using computational methods to look at the distribution of words across a corpus. One set of such tools are distributional semantic models (DSMs). The typical DSM represents terms as geometric vectors in a high-dimensional space that can be compared to give a quantitative measure of similarity. This is typically done by taking the cosine of two vectors, with a value of 1 indicating identical meanings while a cosine of 0 would indicate completely dissimilar meanings.,[18] [19]

There are a number of different ways of carrying out distributional analyses. Unfortunately, the details of these different approaches can get quite daunting, especially for researchers who are new to the area. That said, we believe that even the more accessible techniques for distributional analyses are of value. As such, we encourage experimental philosophers to begin employing these tools, and to tackle their more complex aspects and sophisticated varieties in due course. We will begin with some tools that any experimental philosopher could employ, then expand the analysis to tools that require greater familiarity with programming.

[18] The cosine can also take on a negative value. It is at best unclear how negative values should be interpreted, however, and these are generally treated as being 0.

[19] For more detailed discussions of DSMs see Baroni et al. 2014a, Erk 2012, and Turney and Pantel 2010.

Perhaps the most prominent type of distributional analysis is Latent Semantic Analysis (LSA; Deerwester et al. 1990), and this is the method we will begin with in this section. LSA starts with the texts of a corpus being broken down into predefined documents, such as paragraphs of text. The frequency of each term in the corpus is then counted for each document to produce a term-by-document matrix. It should be noted that this matrix does *not* include information about the relative location of terms in a document. Because of this, LSA is sometimes referred to as a 'bag-of-words' approach. And in this, LSA is perhaps most markedly different from the approach used in the previous section, which specifically looked at the relative position of terms in a sentence.

While LSA has had a good deal of empirical success, one should be mindful of the limitations of the bag-of-words approach and recognize that other approaches are available. In LSA, the context for a target word is the rest of the document. Alternatively, window-based methods use the terms surrounding the target word as context (while this can be thought of as a bag-of-words, it is a relatively *small* bag of words). For instance, a window of size 5 would take the two terms to either side of the target word as context. Another option is to use the words that stand in a particular syntactic relation to the target word as the context.[20] In contrast to these approaches, the 'new kids on the distributional semantics block' are what Baroni et al. (2014b) term *context-predicting models*. Instead of counting the terms occurring in a given context around a target word, these models use artificial neural networks to set vector weights that 'optimally predict the contexts in which the corresponding words tend to appear' (Baroni et al. 2014b, 238).[21]

The easiest way to begin using LSA is to query a premade semantic space. One option is the LSA website from the University of Colorado Boulder.[22] This website allows users to run a number of different types of queries for a range of semantic spaces. To illustrate, we used the pairwise comparison tool for the General Reading up to 1st Year College space[23] to look at cosine values for 'cause' and 'caused' as compared to four terms relevant to assessing causal attributions and their relation to outcome valence – 'responsibility,' 'blame,' 'fault,' and 'praise.'

[20] For a comparison of these approaches, as well as a number of other parameters involved in constructing vector-based DSMs, see Kiela and Clark 2014.

[21] Baroni et al. (2014b) conduct an extensive comparison between context-predicting and count-based models. To their surprise, they 'found that the predict models are so good that [...] there are very good reasons to switch to the new architecture' (245).

[22] See http://lsa.colorado.edu/

[23] This space is built from a corpus of 37,651 documents and covers 92,409 unique terms.

As predicted on the basis of our previous analyses, both terms show a notable similarity to 'responsible'. Further, in line with our previous analyses we found that both terms show a notable similarity with the negative terms 'blame' and 'fault', but essentially no similarity with the positive term 'praise' (see Table 7.3).

It would be nice to be able to say something absolute about the degree of similarity indicated by a given cosine value. Unfortunately, this is complicated by differences in the sizes of LSA spaces. As a result, the values should be thought of as relative measures. One option for getting a sense of the relative values for a space is to test some comparison terms. For instance, the value we found for 'cause' and 'responsible' is slightly higher than the value we find for 'dog' and 'wolf' (0.30), while the value for 'dog' and 'animal' (0.15) is half that, and the value for 'dog' and 'sandwich' is slightly higher than we found for 'cause' and 'praise'. While such comparisons can help you get an initial sense of the space, it is dependent on the terms that you select and might be misleading. A more systematic approach is to look at a predefined list of comparisons. One option is to use a list that is part of a benchmark, such as MEN (Bruni et al. 2013). MEN includes a test set of 1,000 comparisons whose relatedness has been assessed by a large sample of human judges. We can run each of these comparisons, then look at the pairs of terms that have a similar cosine value to the pairs we want to assess.

Another tool available through the LSA website of the University of Colorado Boulder is to search for the nearest neighbours of a given term. This provides the terms closest to the target term *in the semantic space*. When we did this for 'cause' and 'caused', we found that the nearest neighbours, excluding terms sharing the same word stem, generally have a negative cast (e.g. 'damage' (0.66), 'symptoms' (0.61), 'disease' (0.69), 'infections' (0.67)). Again this is in keeping with our previous findings. When we turned to 'responsible', however, we found that many of the nearest neighbours for this term are of a different sort. For instance, we found that 'duties' (0.57) is the nearest neighbour, followed directly by 'supervision' (0.56) and 'personnel' (0.55). This suggests that the responsibility

Table 7.3 Cosine values for term comparisons for the General Reading up to 1st Year College space on the LSA website from the University of Colorado Boulder.

	responsible	**blame**	**fault**	**praise**
cause	0.32	0.28	0.30	0.05
caused	0.32	0.31	0.26	0.01

attributions we are after are getting drowned out by the alternative usage of 'responsibility' noted in the previous section – that of the duties associated with a role.

In addition to getting a sense of degree of similarity indicated by a cosine value in a given space, we also assessed whether it is doing a good job in capturing the semantic relatedness of terms. To do this we employed the MEN list, mentioned above, and used the list of cosine values for the test set to analyse how well this correlates with the scores from the human judges. When we did this for the General Reading space, we found that it does a relatively good job: we got a Spearman's rho of 0.67. For comparison, Kiela and Clark (2014) report values of 0.66 to 0.71 for the spaces they compared in their Table 6, while Baroni et al. (2014b) report in Table 7.2 a top value of 0.72 for the best count-based model tested and a value of 0.80 for the best context-predicting model tested.

It is also possible to build an LSA space oneself. While the details that go into the construction of an LSA space are complicated, a number of tools are available to facilitate the process. We will begin with tools that can be used through the statistical software package R. While there are several benefits to using R in the present context,[24] it also suffers from some limitations, as we will see.

The easiest way to get started with LSA in R is to use the LSAfun package to import a premade semantic space (Günther et al. 2015). To illustrate, we used the EN_100k_lsa space to further explore the relationship between 'cause'/'caused' and 'responsible'. This space was created from a corpus of some two billion words combining the British National Corpus, the ukWaC corpus and a 2009 Wikipedia dump. We began by looking at the same set of comparisons that we performed above. Again we see that 'cause'/'caused' show a notable similarity to 'responsible', and that both terms are much more similar to 'blame' and 'fault' than to 'praise' (see Table 7.4). Next, we looked at the nearest neighbours for 'cause', 'caused' and 'responsible'. The results were in line with what we saw previously, with 'cause' and 'caused' being close to a number of negative terms (e.g. 'excessive' (0.78), 'suffer' (0.78), 'damage' (0.82), 'fatal' (0.71)), while 'responsible' was close to a range of terms related to duties associated with a role (e.g. 'overseeing' (0.76),

[24] One is that R supports a large range of statistical analyses of use to experimental philosophers. Another is that the only book-length guide to the practice of experimental philosophy currently available (Sytsma and Livengood 2016) uses R as its preferred statistical program and Chapter 10 of that work provides a general introduction to the use of R.

Table 7.4 Cosine values for term comparisons for the EN_100k_lsa space.

	responsible	**blame**	**fault**	**praise**
cause	0.32	0.46	0.50	0.23
caused	0.34	0.47	0.57	0.18

'supervising' (0.63)). Like the previous space, the EN_100k_lsa space performs well on the MEN benchmark with a Spearman's rho of 0.67.[25]

R also offers tools for creating corpora and building LSA spaces. To illustrate, we used the RCurl package to scrape the text for all entries in the Stanford Encyclopedia of Philosophy and the Internet Encyclopedia of Philosophy from their websites. The result was a corpus including 2,378 entries split into 136,946 paragraphs (the documents for our analyses) and composed of over 149 million words and 115,644 unique terms. We then used the koRpus package, to annotate the documents with lemma information. The tm package was used to convert this into a corpus, which was fed into the lsa package to generate the term-by-document matrix and create the semantic space. It is in this final step that we ran into the limitations of R noted above. Specifically, in R the data for analysis is held in RAM, which places severe limits on the size of the matrix it can process on a typical home computer. One option is to reduce the size of the term-by-document matrix by removing infrequently occurring terms before creating the semantic space. For the philosophy corpus, we needed to reduce the matrix to the 8,821 most frequently occurring terms.

Given that philosophers have often treated the ordinary concept of causation as being a purely descriptive concept and that many have expressed either surprise or outright scepticism towards the results surveyed above we would not expect to find the valence effect for the philosophy corpus that we saw in our previous investigations. With regard to the relation between 'cause' and 'responsible', one might predict that these terms would also be largely unrelated. Alternatively, one might note that many philosophers hold that causing an outcome is a prerequisite for being responsible for that outcome. As such, one

[25] There is also a window-based space available from the same corpus – the EN_100k space – that performs slightly better on the MEN benchmark (Spearman's rho of 0.71). Another option are the spaces available from the COMPOSES Semantic Vectors website, which provides the best models from Baroni et al. (2014b) as text files. Their context-predicting model performs especially well with a Spearman's rho of 0.80 on the MEN benchmark. Both spaces paint a similar picture to what we saw for the EN_100k_lsa space, both in terms of the comparisons in Table 7.4 and the nearest neighbours of 'cause'/'caused' and 'responsible'.

might expect to see a notable similarity between these terms in the philosophy corpus. What we found is that 'cause' showed virtually no relation to 'blame', 'fault' or 'praise', and that it showed virtually no relation to 'responsible' (see Table 7.5).[26] The space performed better than expected on the MEN benchmark, with a Spearman's rho of 0.48. Because the term-by-document matrix was significantly reduced, however, the correlation was only calculated on 269 comparisons.

Given the degree to which the term-document-matrix was reduced, the results for the philosophy corpus LSA space should be taken with a hefty grain of salt. To further test these results, we switched to the Gensim toolkit implemented in Python, which is designed to handle large corpora efficiently and is able to carry out a wide variety of distributional analyses, including LSA and the context-predicting models using word2vec that performed best in Baroni et al.'s (2014b) comparisons. We exported the processed philosophy corpus from R into Gensim, then analysed it using word2vec with recommended parameters, including a five-word context window. The results were quite different from what we found for the LSA space. Most notably, we found a much higher cosine value for 'cause' and 'responsible' (see Table 7.6). We also saw a moderate relation between 'cause' and 'blame' or 'fault', but no relation between 'cause' and 'praise'. Further, the nearest neighbours of 'cause' and 'responsible' were quite distinct.[27] While these results are more like what we've seen for the general corpora, they continue to

Table 7.5 Cosine values for term comparisons for the philosophy corpus LSA space.

	responsible	**blame**	**fault**	**praise**
cause	–0.0003	0.0004	–0.0011	–0.0051

Table 7.6 Cosine values for term comparisons for the philosophy corpus word2vec space.

	responsible	**blame**	**Fault**	**praise**
cause	0.41	0.18	0.17	–0.04

[26] We ran the comparisons only for 'cause' since we lemmatized the text and 'cause' and 'caused' belong to the same lemma.

[27] Five nearest neighbours for 'cause': 'proximate' (0.64), 'efficient' (0.60), 'effect' (0.59), 'volition' (0.51) and 'necessitate' (0.50); five nearest neighbours for 'responsible': 'accountable' (0.69), 'blameworthy' (0.64), 'attributable' (0.58), 'culpable' (0.56) and 'negligent' (0.55).

suggest that the philosophical usage diverges from the ordinary usage, as will be spelled out below. The space performed comparably on the MEN benchmark, with a Spearman's rho of 0.48 on a much higher number of comparisons.[28]

The same tools can be applied to other corpora that are available for download, including COCA. To facilitate comparison to philosophical usage, we excluded academic texts. Since COCA comes with lemma information, we did not need to annotate the documents. Other than this we followed the same procedure detailed for the philosophy corpus to generate a word2vec space. The results were in keeping with what we saw for the other general corpora above, with there being a notable similarity between 'cause' and 'responsible', between 'cause' and the negative lemmas 'blame' and 'fault', and no similarity between 'cause' and the positive lemma 'praise' (see Table 7.7). As expected, the space performed extremely well on the MEN benchmark with a Spearman's rho of 0.80.

With access to a full corpus it is also possible to target causal attributions and responsibility attributions more carefully by directly comparing multi-word expressions. To do this we replaced the phrases 'caused the' and 'responsible for the' with single tokens before lemmatizing and processing the non-academic COCA corpus. We then analysed it using the same predictive model as above. The cosine value between the causal attribution token and the responsibility token was quite large, and notably larger than for the previous comparison between 'cause' and 'responsible' (see Table 7.8). Further, each token was one of

Table 7.7 Cosine values for term comparisons for the non-academic COCA corpus word2vec space.

	responsible	**blame**	**fault**	**praise**
cause	0.41	0.56	0.48	–0.06

Table 7.8 Cosine values for term comparisons for the non-academic COCA corpus word2vec space with causal attribution and responsibility attribution tokens.

	responsible for the	**blame**	**fault**	**praise**
caused the	0.63	0.55	0.42	–0.10
responsible for the		0.61	0.35	0.16

[28] 944 of the 1,000 comparisons were used (25 terms were missing from the corpus).

the five nearest neighbours of the other. The nearest neighbours for each token included a number of terms with a negative cast – e.g. ‘catastrophic’ (0.68), ‘fatal’ (0.63), ‘culpable’ (0.62), ‘complicit’ (0.58) – including that ‘blame’ was one of the five nearest neighbours for the responsibility attribution token. In addition, none of the nearest neighbours for this token indicated the notion of duties associated with a role that marred our previous results.

To better compare philosophical usage with ordinary usage, we tokenized the philosophy corpus and repeated the analysis. We found that the causal attribution token was quite similar to the responsibility attribution token. In line with the alternative prediction noted above, this might reflect that many philosophers hold that causing an outcome is a prerequisite for being responsible for that outcome. Although the cosine values for the two tokens are similar for both the philosophy space and the COCA space, when we look deeper we find evidence that the causal attributions of philosophers are quite different from the causal attributions of more ordinary language. Thus, while there was a strong relation between the two tokens in the philosophy space, the causal attribution token was much less similar to ‘blame’ and ‘fault’ (see Table 7.9). This stands in marked contrast to what we saw for the COCA space. Further, the same contrast holds for responsibility attributions. The results suggest that the ordinary usage of both causal attributions and responsibility attributions has a negative cast that the philosophical usage lacks.

Overall the results of our latent semantic analyses nicely line up with the results of the analyses in the previous section, with the two approaches providing a consilience of evidence. Looking across the two sets of analyses, we find compelling evidence for a positive answer to each of the questions we opened this chapter with: our results provide independent support for the thesis that ordinary causal attributions are sensitive to normative information; they provide support for the contention that outcome valence often matters for ordinary causal attributions; they indicate that causal attributions are similar to responsibility attributions[29]; and they suggest that philosophers use the language of causal attribution differently from lay people.

[29] It could be objected here that while our results indicate that causal attributions are similar to responsibility attributions, they do not indicate that causal attributions are themselves normative. For instance, it might be suggested that they are close in semantic space because causing an outcome is a prerequisite for being responsible for that outcome. This would not explain the results of our analysis in the previous section, however, or the valence effect observed for causal attributions in the semantic spaces. While this could be investigated further using DSMs to assess the Synonymy Effect hypothesis, space prevents us from doing so here. Alternatively, expanding on Alicke’s view, it might be argued that the desire to blame biases both causal attributions *and* responsibility attributions. While we cannot rule this out based on the present results, we hold that the responsibility view offers the simpler explanation.

Table 7.9 Cosine values for term comparisons for the philosophy corpus word2vec space with causal attribution and responsibility attribution tokens.

	responsible for the	**blame**	**fault**	**praise**
caused the	0.55	0.16	0.17	0.03
responsible for the		0.10	0.07	0.00

6. Concluding remarks: Corpus analysis as a method for experimental philosophy

Causation is one of the most contested concepts in philosophy. Recent questionnaire-based studies have produced some rather surprising insights into how we use that concept. Most importantly, they suggest that normative considerations play a central role in ordinary causal attributions. However, it is still an open issue how best to interpret and explain these results.

We believe that to make progress in deciding between the different accounts of the impact of norms on causal attributions, it is fruitful to expand the set of empirical tools used by experimental philosophers working on causation. Specifically, we believe that the study of ordinary causal attributions can benefit from the tools of corpus linguistics. One reason for this belief is that while questionnaire methods are powerful and often well-suited to investigating philosophical questions, they also have limitations. And the questionnaire-based studies on ordinary causal attributions that we have looked at in this chapter do suffer from some of those limitations. To illustrate, we will focus on the first study we discussed in Section 2 Knobe and Fraser's (2008) investigation of the Pen Case.[30]

The basic misgiving one might have about Knobe and Fraser's study is that it relies on an instrument they unintentionally designed in such a way that it would elicit the suspected effect, not because norms actually do have an impact on ordinary causal attributions but because the study leads participants to misread the prompts. For instance, one might object to participants being asked which one of the two agents, the administrative assistant or the professor, caused the problem. While this is certainly a very natural way to ask the question of interest, we believe that asking people who caused 'the problem', as compared

[30] We would like to emphasize that even though we focus on Knobe and Fraser here, our worries apply to questionnaire studies more generally, and the authors recognize that their own work is also liable to these methodological concerns.

to 'the outcome' or 'the situation', might trigger an interpretation of the question in normative terms. Alternatively, one might worry that by having participants rate two statements – one about the administrative assistant and one about Professor Smith – and phrasing this in a way that suggests an all-or-nothing state of affairs (as opposed, for example, to asking whether the agent was 'a cause' of the outcome) might prompt participants to feel that they should agree with at most one of the two questions. Since the only distinguishing feature between the two agents' actions is that one violates a norm while the other does not, participants might latch onto this as a relevant cue for fulfilling the task. Or one might note that Knobe and Fraser asked participants whether an *agent* caused an outcome. Typically, when philosophers talk about causation, they talk about *events* as the causal relata, not *people*. Asking about the agent, rather than his or her activities, might therefore create another reason for participants to believe that the researcher is asking about something normative.

All of the potential issues we just noted for Knobe and Fraser's study could be addressed through further questionnaire-based research. And, in fact, a good deal of work has subsequently been done on the Pen Case, or cases like it, that varies these sorts of factors. But follow-up studies addressing one potential confound run the risk of introducing others. This is simply one of the difficulties inherent to this type of research. It does not mean, of course, that questionnaire studies should be abandoned. Rather, the moral we should draw from it is that in the face of these risks we should diversify our set of methods. Turning to corpus linguistics seems natural here, as one strength of corpus analysis is that it is relatively immune from the pragmatic pitfalls we have just highlighted.

One of the motivations of corpus linguistics is the preference for 'real' language data over examples of language use generated by linguists themselves. While a corpus cannot, strictly speaking, be representative for a language in its entirety (because possible utterances of that language are infinite), linguistic corpora aim at balanced sampling from this impressive population. Unless the research interest is focused on a specific genre – such as the usage of a given term in academic texts – a balanced corpus contains a considered choice of texts of various types and from different sources. For example, it does not only contain written, but also (transcribed) spoken language, not only specialized (e.g. academic) language, but also its everyday variety, not only literary texts, but also mundane ones such as operation manuals. In this way, a large corpus does present a meaningful sample of actual language use.

The preference for 'real' language sits nicely with precepts of experimental philosophy, in that it emphasizes the importance of empirical data over that generated by researchers relying on their own intuition or judgement. An important advantage it has over data generated by questionnaire studies is that the linguistic data of a general corpus usually has been generated independently of the researcher and her specific research questions. The data thus is usually uncued, in the sense that the utterances the corpus contains have not been produced in response to some prompt of the researcher. It is then plausible to assume that a corpus is unbiased with respect to the specific research question with which a philosopher approaches it (cf. Schütze 2010).[31] And for the same reason, such a corpus can be considered free of the biases of experimental pragmatics.

Having said that, there are, of course, limits to corpus analysis. The linguistic data of a corpus can be evidence in relation to some philosophical issue only to the extent that the actual use of language is relevant to it. This relevance may be direct or indirect, because the linguistic data in a corpus may well allow us to infer something about deeper structures of language. But, clearly, if the observation of linguistic phenomena in actual use is irrelevant to a philosophical issue, then so is corpus analysis. Moreover, if the pertinent phenomena are of the very particular and subtle kind that is common for philosophical problems, even a large corpus may not yield any, let alone many, examples of their use. By way of contrast, questionnaires can be constructed to elicit informants' responses to precisely worded prompts, and in doing so, the wording can easily be varied to bring out and test subtle differences in language.

Therefore it is clear that corpus analysis is best viewed as a fruitful addition to the methodological toolbox of experimental philosophy. Not only can it be used effectively to explore the actual use of linguistic expressions – something that is called for in philosophy on many occasions – it can, more specifically, be used to complement experimental studies in several helpful ways: to pre-test hypotheses that inform such studies, to help with the general construction of questionnaires and the precise wording of their items, and, most importantly, we believe, to test the findings from questionnaire studies, either giving them independent support from another empirical source or providing evidence against them.

[31] It is, of course, possible to come up with examples of corpora that are dependent on the researcher and that contain cued language use. Most simply, for example, in the case that the corpus consists of written responses to a vignette. It is the choice of texts that determines whether the data contained in a corpus is indeed independent and uncued. If a pre-existing general language corpus is used, this objection can be assumed to be moot.

Suggested readings

Bluhm, R. (2016). Corpus analysis in philosophy. In M. Hinton (ed.), *Evidence, Experiment and Argument in Linguistics and the Philosophy of Language* (pp. 91–109). Frankfurt am Main: Peter Lang.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, *6*(10), 635–653.

McEnery, T., and Wilson, A. (2001). *Corpus linguistics* (2nd edn.). Edinburgh: Edinburgh University Press.

Turney, Peter, and Pantel, Patrick (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.

Bibliography

Achugar, M, and Schleppegrell, M. J. (2005). Beyond connectors. *Linguistics and Education*, *16*, 298–318.

Alicke, M. D. (1992). Culpable causation. *Journal of Personality and Social Psychology*, *63*, 368–378.

Altenberg, B. (1984). Causal linking in spoken and written english. *Studia Linguistica*, *38*, 20–69.

Baroni, M., Bernardi, R., and Zamparelli, R. (2014a). Frege in space. *Linguistic Issues in Language Technology*, *9*, 241–346.

Baroni, M., Dinu, G., and Kruszewski, G. (2014b). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 238–247.

Bergh, G., and Zanchetta, E. (2008). Web linguistics. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics*, Vol. 1 (pp. 309–327). Berlin: de Gruyter.

Biber, D. (2010). Corpus-based and corpus–driven analyses of language variation and use. In B. Heine and H. Narrog (eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 159–191). Oxford: Oxford University Press.

Biber, D., Conrad, S. and Reppen, R. (1998). *Corpus Linguistics*, Cambridge: Cambridge University Press.

Bluhm, R. (2012). *Selbsttäuscherische Hoffnung: Eine sprachanalytische Annäherung*. Münster: mentis

Bluhm, R. (2013). Don't ask, look! Linguistic corpora as a tool for conceptual analysis. In M. Hoeltje, T. Spitzley, and W. Spohn (eds.), *Was dürfen wir glauben? Was sollen wir tun? Sektionsbeiträge des achten internationalen Kongresses der Gesellschaft für Analytische Philosophie e.V*. Duisburg: DuEPublico.

Bluhm, R. (2016). Corpus analysis in philosophy. In M. Hinton (ed.), *Evidence, Experiment and Argument in Linguistics and the Philosophy of Language* (pp. 91–109). Frankfurt am Main: Peter Lang.

Bruni, E., Tran, N. K. and Baroni, M. (2013). Multimodal distributional semantics. *Journal of Artificial Intelligence Research 49*, 1–47.

Cushman, F. (2013). Outcome, and value: A dual-system framework for morality. *Personality and Social Psychology Review*, *17*(3), pp. 273–92.

De Schryver, G.–M. (2002). Web for/as corpus. *Nordic Journal of African Studies*, *11*, 266–282.

Deerwester, S., Dumais, S., Landauer, T., Furnas, G., and Harshman, R. (1990). Indexing by Latent Semantic Analysis. *Journal of the Society for Information Science*, *41*(6), 391–407.

Diessel, H., and Hetterle, K. (2011). Causal clauses: A crosslinguistic investigation of their structure, meaning, and use. In P. Siemund (ed.), *Linguistic Universals and Language Variation* (pp. 23–54). Berlin: Mouton de Gruyter.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A survey. *Language and Linguistics Compass*, *6*(10), 635–653.

Firth, J. R. (1957). A synopsis of linguistic theory 1930–55. In *Studies in Linguistic Analysis* (pp. 1–32). Oxford: Blackwell.

Fischer, E., and Engelhardt, P. E. (2017). Diagnostic experimental philosophy. *teorema*, *36*(3), 117–137.

Günther, F., Dudschig, C., and Kaup, B. (2015). LSAfun: An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*, 930–944.

Hahn, U., Zenker, F., and Bluhm, R. (2017). Causal argument. In M. R. Waldmann (ed.), *The Oxford Handbook of Causal Reasoning* (pp. 475–494). New York, NY: Oxford University Press.

Harris, Z. (1954). Distributional structure. *Word*, *10*(23), 146–162.

Henne, P., Pinillos, Á., and Brigard, F. De (2015). Cause by omission and norm: Not watering plants. *Australasian Journal of Philosophy*, *XXX*, 1–14.

Hilton, D., and Slugoski, B. (1986). Knowledge-based causal attribution: The abnormal conditions focus model. *Psychological Review*, *93*, 75–88.

Hitchcock, C., and Knobe, J. (2009). Cause and norm. *The Journal of Philosophy*, *106*, 587–612.

Khoo, C., Chan, S., and Niu, Y. (2002). The many facets of the cause-effect relation. In R. Green, C. A. Bean, and S. H. Myaeng (eds.), *The Semantics of Relationships* (pp. 51–70). Dordrecht: Springer Netherlands.

Kiela, D., and Clark, S. (2014). A systematic study of semantic vector space model parameters. *Proceedings of the 2nd Workshop on Continuous Vector Space Models and their Compositionality*, 21–30.

Kilgarriff, A., and Greffenstette, G. (2003). 'Introduction' to 'the Web as Corpus'. *Computational Linguistics*, *29*(3), 333–347.

Knobe, J., and Fraser, B. (2008). Causal judgments and moral judgment: Two experiments. In W. Sinnott–Armstrong (ed.), *Moral Psychology, Volume 2: The Cognitive Science of Morality* (pp. 441–447). Cambridge: MIT Press.

Kominsky, J., Phillips J., Gerstenberg, T., Lagnado, D., and Knobe, J. (2015). Causal superseding. *Cognition*, *137*, 196–209.

Lee, D. Y. W. (2010). What corpora are available? In M. McCarthy and A. O'Keeffe (eds.), *Corpus Linguistics* (pp. 107–121). London, New York: Routledge.

Leech, G. (1992). Corpora and theories of linguistic performance. In J. Svartvik (ed.), *Directions in Corpus Linguistics* (pp. 105–122). Berlin, New York: Mouton de Gruyter.

Livengood, J., Sytsma J., and Rose, D. (2017). Following the FAD: Folk attributions and theories of actual causation. *Review of Philosophy and Psychology*, *8*(2), 274–294.

Livengood, J., and Sytsma, J. (under review). *Actual Causation and Compositionality*.

Lüdeling, A., and M. Kytö (eds.) (2008–2009). *Corpus Linguistics* (2 vols.). Berlin: de Gruyter.

Malle, B., Guglielmo S., and Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry 25*(2), 147–86.

McCarthy, M., and O'Keefe, A. (2010a). Historical perspective. In M. McCarthy and A. O'Keefe (eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 3–13). London, New York: Routledge.

McCarthy, M., and O'Keefe, A. (eds.) (2010b). *The Routledge Handbook of Corpus Linguistics*. London, New York: Routledge.

McEnery, T., and Wilson, A. (2001). *Corpus linguistics* (2nd edn.). Edinburgh: Edinburgh University Press.

Prinz, J. (2007). *The Emotional Construction of Morals*. Oxford: Oxford University Press.

Reuter, K. (2011). Distinguishing the appearance from the reality of pain. *Journal of Consciousness Studies*, *18*(9–10), 94–109.

Reuter, K., Kirfel, L., Riel, R. van, and Barlassina, L. (2014). The good, the bad, and the timely: How temporal order and moral judgment influence causal selection. *Frontiers in Psychology*, *5*, 1336.

Samland, J., and Waldmann, M. R. (2014). Do social norms influence causal inferences? In P. Bello, M. Guarini, M. McShane, and B. Scassellati (eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 1359–1364). Austin, TX: Cognitive Science Society.

Samland, J. and Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, *156*, 164–176.

Schütze, C. T. (2010). Data and evidence. In K. Brown, A. Barber, and R. J. Stainton (eds.), *Concise Encyclopedia of Philosophy of Language and Linguistics* (pp. 117–123). Amsterdam et al.: Elsevier.

Sytsma, J., Livengood, J., and Rose, D. (2012). Two types of typicality: Rethinking the role of statistical typicality in ordinary causal attributions. *Studies in History and Philosophy of Biological and Biomedical Sciences*, *43*, 814–820.

Sytsma, J., Rose, D., and Livengood, J. (ms). *The Extent of Causal Superseding*.
Sytsma, J., and Livengood, J. (2016). *The Theory and Practice of Experimental Philosophy*. Peterborough: Broadview.
Sytsma, J., and Livengood, J. (2018). *Intervention, Bias, Responsibility … and the Trolley Problem*. http://philsci-archive.pitt.edu/14549/.
Sytsma, J., and Reuter, K. (2017). Experimental philosophy of pain. *Indian Council of Philosophical Research*, *34*(3), 611–628.
Turney, Peter and Pantel, Patrick (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188.
Willemsen, P. (2016). Omissions and expectations: A new approach to the things we failed to to do. *Synthese*, *19*(4), 1587–1614.
Willemsen, P. and Reuter, K. (2016). Is there really an omission effect? *Philosophical Psychology*, *29*(8), 1142–1159.
Xiao, R. (2008). Well-known and influential corpora. In A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics*, Vol. 1 (pp. 383–457). Berlin, New York: de Gruyter.