Corpus-linguistic approaches to metaphor analysis

Heli Tissari

Introduction

Corpora should be used to study metaphors for the same reasons that they should be used for any linguistic research. They help us to search through a lot of data with great speed, they make research replicable, and they give us access to statistics on linguistic phenomena. This chapter will discuss and demonstrate how best to make use of corpora in metaphor research.

What is a corpus and why should we use one to study metaphors?

A corpus is a collection of texts, and an electronic corpus is a computer-readable corpus, which, when fed into appropriate software, makes searching and comparing different texts easy and fast. You can search corpora in many different ways. You can search them for words, morphemes, colligations, collocations and syntactic structures. Moreover, many corpora come with so-called annotation, i.e. they contain information about the syntactic or semantic features of their contents. Colligation refers to a grammatical combination of two or more words. For example, the verb agree tends to go together with the preposition to. Collocation refers to the habitual co-occurrence of words. For example, the adjective auspicious tends to go together with such nouns as occasion and event (Crystal 2008: 86).

Corpus-based research becomes particularly replicable if you use a publicly available corpus, such as the 100-million-word British National Corpus BNC or the 450-million-word Corpus of Contemporary American English COCA (Davies 1990–2012). A corpus interface often gives us direct access to statistics. For example, the BNCWeb (CQP edition) (Hoffmann 1996–2013) tells us which words we should expect to go together and which actually do.

When using a corpus, it is important to know what kind of texts it consists of. If you search a corpus for something, you need to know which variety, genre and topic the samples represent, and then you can count likelihoods for the occurrences of certain phenomena in certain types of text. For example, the Brown Corpus was created in the 1960s by collecting samples of many different genres in order to represent then current American English, and the Corpus of Early English Correspondence Sampler (1418–1680) represents letters. Moreover, we can distinguish between small and big corpora. The one-million-word
Brown Corpus is small while, for example, the BNC is big, not to mention the 650-million-word Bank of English (see The Collins Corpus 2007–2013). The bigger a corpus is, the more likely one is going to find rare words in it.

Corpus studies on metaphor can be divided into three types (Semino 2008: 199). In the first type, researchers use general-purpose corpora such as the Brown Corpus to search for patterns of conceptual metaphors which could be relevant for the development of Conceptual Metaphor Theory (CMT; see Chapter 1). In the second type, researchers choose corpora representing text types, genres or particular periods, in order to investigate similarities and differences between these. In the third type, researchers compare metaphors occurring in corpora representing different languages (examples of all these types of studies will be given below). Bigger corpora would seem to lend themselves better to metaphor research than small ones, because they contain more data even on rare expressions. However, sometimes the amount of data in a big corpus is too much. Koller (2006: 242) notes that large amounts of corpus can be an obstacle for meticulous metaphor analysis. However, large and small corpora can also be combined in research (e.g. Cameron and Deignan 2003).

In the following sections, I will provide a sample method for how corpora can be used in metaphor research and outline previous studies where this has been done. I will then move on to demonstrate what can be gained from a corpus approach to metaphor analysis with a case study of recent research. The final section will briefly outline the development of corpus linguistic research on metaphor in the past and how I expect it to develop in the future.

How to use corpora to study metaphor: in a nutshell

To use corpora to study metaphor, you need to consider the following questions (not necessarily in this order):

1. What is your topic (e.g. metaphors realized by verbs, metaphors of emotion)?
2. Which corpus or corpora will you use as data? Why?
3. How does your research relate to previous studies?
4. How do you define metaphor?
5. How are you going to identify the metaphors in the data? (What are you going to search for in the corpus and how are you going to process the search results?)

Overview of previous research

A preliminary note and a definition of metaphor

My overview of previous research mostly relates to CMT as developed by Lakoff and Johnson (e.g. 1980, 1999). Many pioneering studies on conceptual metaphors from the 1980s and 1990s did not use corpora as data (e.g. Kövecses 1986, 1990; Lakoff 1987: 380–415). We need not dismiss those studies even if we embrace corpora. On the contrary, we can use corpora to check theories and hypotheses which have been created without them (e.g. Deignan 2005; Stefanowitsch 2006b).

According to CMT, a metaphor consists of a source domain and a target domain. The target domain is the concept which we understand in terms of the other concept, the source domain (Lakoff 1987: 384). For example, the conceptual metaphor love is a physical force consists of the source domain physical force and the target domain love. It may be realized in the form of example (1) (Lakoff and Johnson 1980: 49):
(1) I could feel the electricity between us.

We could find this metaphor in a corpus by searching for the source domain word electricity, but not necessarily by searching for the target domain word love. To be precise, CMT suggests that metaphors exist primarily in the mind, although they are expressed in language.

Examples of previous research

Deignan (e.g. 1999, 2005, 2008) showed an early interest in combining metaphor research with corpus linguistics. She was both supportive and critical of CMT. One of her major aims was to show that authentic corpus data behaves differently from data gathered by introspection. She was interested, for example, in how frequently different parts of speech occur with a metaphorical sense. She suggested that, for example, animal metaphors such as Richard is a gorilla where the metaphorical word, gorilla, is a noun, are rare. On the other hand, many animal metaphors occur in the form of verbs (e.g. to horse) or adjectives (e.g. sheepish) (Deignan 2005: 152–5). She also studied colligation and collocation, including the words occurring together with the noun price, and noticed that some colligates, such as the preposition of in price of, primarily represented literal meanings, while some collocates, such as heavy in a heavy price to pay, primarily represented metaphorical meanings (Deignan 2005: 203–9).

Recent innovative corpus-linguistic work on metaphors includes Philip’s (2011) research on idioms and collocations. She studied such expressions as caught red-handed, once in a blue moon, red tape and the grass is always greener, discovering, for example, that although red-handed in caught red-handed would seem to refer to violence, people typically caught red-handed in a crime have been ‘dealing or using drugs, smuggling arms, stealing, [or] committing fraud’ (Philip 2011: 91). As regards the expression the grass is still green, Philip (2011: 146–8) searched for variant forms noting, for example, that people never said the grass is still green, the grass is usually green or the grass tends to be green, and that they replaced the adjective green only once with another adjective denoting colour, i.e. black.

You can also use corpora to compare different languages with each other. Ureña Gómez-Moreno and Faber (2011) studied metaphorical terminology in a corpus of marine biology journals written in English and Spanish. They searched for species names based on resemblance metaphors, such as harvest fish and hammerhead shark. They found three kinds of metaphorical terms: exact pairs, in which the English and Spanish terms were based on the same metaphors, separate pairs, in which the English and Spanish terms were based on different metaphors, and unbalanced pairs, in which only one language used a metaphor.

The corpus-linguistic method has also been used to complement manual analyses. Hardie et al. (2007) and Koller et al. (2008: 153–6) studied the metaphorical source domains in Ken Kesey’s novel One Flew over the Cuckoo’s Nest. They established that Semino and Swindlehurst’s (1996) manual analysis of the machinery metaphors in the novel was correct and that the metaphors people are machines and institutions are machines often occurred in the first half of the text. Hardie et al. (2007) also compared Koller’s (2004b) manual analysis of 40 different articles on 20 different businesswomen with a computerized analysis and found that the latter method yielded ‘two to three times more results’.

Sometimes we find a series of articles on a topic, such as emotion. The seminal study was by Lakoff and Kövecses on the American English concept of anger (Lakoff 1987: 380–415). Kövecses continued working on emotions, publishing, among other things, a book called...
Emotion Concepts (1990), where he defined a prototype of emotion on the basis of metaphors of emotion. Corpus-based studies have also investigated metaphors for target domains that received a lot of attention in the CMT literature, such as emotions (e.g. Lakoff 1987; Kövecses 2000). For example, Stefanowitsch (2006b) searched for some emotion words in the BNC and identified some metaphors that had not been noted in the literature. Tissari (2003) first continued Kövecses’s (1988) work on love and then turned to other emotions, such as pride (Tissari 2006b). Her work on pride in Late Middle (1418–1500) and Early Modern English (1500–1710) was followed by Fabiszak and Hebda’s (2010) work on pride in Old (–1150) and Middle English (1150–1420).

How to find metaphors in corpora

The definition of conceptual metaphor suggests two ways to find metaphors in corpora. You can search for linguistic items which relate to the source domain or search for linguistic items which relate to the target domain (Stefanowitsch 2006a: 2–4). Charteris-Black (2004) searched for source domain lexis when he studied e.g. the metaphors of New Labour, metaphor in British party political manifestos and metaphor in sports reporting. He first read a sample of a certain text type in order to identify expressions suggesting metaphor sources, and then searched for similar expressions in a larger corpus. For example, he read a ‘sample of 100 sports reports from popular and broadsheet newspapers’ to compile a lexicon listing conflict terms which could be used to metaphorically suggest that sport is war (Charteris-Black 2004: 117). The list included such words as war, attack, fight, kill, campaign, army, battle, victory, victim and struggle. He then searched for these words in The Times and The Sun newspapers to see how often sports was characterized as war (Charteris-Black 2004: 116–17). Koller (2004a) and Hintikka (2013) used variants of this approach in order to study metaphor and gender in business media discourse and body metaphors for society and the mind in four corpora representing Early Modern and Present-Day English. The reverse method of searching for target domain lexis and identifying metaphors occurring with it has been pioneered by Stefanowitsch (2006b) and Tissari (2003). It will be described below.

To find metaphors in a corpus, you can also search for so-called tuning devices of metaphors, i.e. words which can be used to introduce metaphors. Cameron and Deignan (2003: 152) used a small corpus consisting of talk recorded in a primary school to identify such devices and came up with the set: actually, almost, imagine, just, kind of, a little, really, sort of. These words were found to serve two functions in teachers’ speech: they directed the pupils to a particular interpretation, and they adjusted the strength of the metaphor, as in the following example which related to dancing practice:

(2) . . . can you go back just a “whisper.”

Cameron and Deignan 2003: 153

Cameron and Deignan (2003: 154–9) then searched for the same tuning devices in a large corpus, finding more uses to which the tuning devices were put: they were again used to direct the interpretation of metaphors and to adjust the strength of the metaphor, but they were also used to signal something unexpected like “midwife” in example (3) (Cameron and Deignan 2003: 156):

(3) Fryer he was the he was er in a way our “midwife” because he was Secretary of the Agricultural Research Council.
Berber Sardinha (2012) compared the success rates of different types of searches for metaphors in corpus data. First, he tested the method of reading samples of text and then searching for the same metaphors in the entire corpus. His findings suggested that you should read at least 30 per cent of a corpus, which, in his test, produced 55.2 per cent of the metaphors in the entire data (Berber Sardinha 2012: 28–31). Second, he tested searches with single words and different combinations of words such as two-word bundles and five words to the left of a chosen node word. He found that searches for two-word bundles, three-word bundles and four-word bundles attested the most precision in identifying metaphors, 100 per cent, while single word searches attested the least precision, 73.2 per cent (Berber Sardinha 2012: 31–4). Third, he experimented with metaphor clustering, testing different windows within which to find more metaphors around a metaphor. He naturally found more metaphors when he enlarged the window size from 5 to 20 words. However, even a 20-word window size, which almost meant reading through the corpus, did not yield all the metaphors there. Consequently, he recommended a 5- or 10-word window size, which would be likely to retrieve 20 per cent to 40 per cent of the metaphors (Berber Sardinha 2012: 34–6). In contrast, Berber Sardinha (2012: 36–8) found that a keyword analysis using a standard corpus comparison tool is not an efficient way to find metaphors.

Until now, we have dealt with searches which can be conducted by using basic corpus tools. The next section will introduce more computational approaches, which combine several tools and/or require programming skills.

**How to find metaphors in the data: going more computational**

Early attempts to program metaphor detection and identification included Martin (1990) and Fass (1991). Both involved a previously programmed knowledge representation system and had some measure of success. More recently, Mason (2004) proposed a different approach based on a simple idea: to compare different text types representing potential source and target domain concepts such as laboratory and finance in terms of their frequent lexis, and then see where there was overflow of lexis from one domain to another. He used the Internet as a source of data. Mason (2004: 29) detected an asymmetric structure of transfer between the domains laboratory and finance: verbs which often occurred with liquids in the laboratory domain, for example, pour, flow and evaporate, also occurred with money in the finance domain, but the reverse did not happen: verbs which often occurred with money, for example, spend, invest and deposit, did not occur with liquids. This is how Mason (2004: 35) detected the metaphor money is a liquid. His finding corroborated the unidirectionality of metaphor as postulated in CMT: the relationship between the source and target domains cannot be reversed (e.g. Sweetser 1990: 30). Mason’s approach (2004) did not require any previously programmed knowledge representation system, but he did use the WordNet lexical database for English as a knowledge base (Fellbaum 2006).

In an alternative approach, Berber Sardinha (2012: 44–7) tested the idea that metaphors could be found by locating two semantically unrelated words near each other. He used software called WordNet::Similarity to identify word pairs which were semantically distant from each other (cf. Pedersen 2014). The words in his data had so many different senses in the WordNet lexical database that the number of word pairs studied multiplied from 12,055 to 343,347. The top 1,000 word pairs produced only seven unique metaphors, although the precision was 100 per cent. He eventually suggested that a 15 per cent sample of the output would be ideal, revealing about 85 per cent of the metaphors in the data (Berber Sardinha 2012: 44–7).
Reining and Lönneker-Rodman (2007) introduced corpus-driven metaphor harvesting. They combined a search for a target domain word, Europe, with searches for source domain lexis in a corpus of articles on the European Constitution in the French newspapers Le Monde and Figaro. They conducted a statistical analysis to see which hundred collocates typically occurred with the word Europe in each newspaper. They then evaluated the top collocates to see which lemmas could be used metaphorically and came up with such words as construire ‘construct’ and traverser ‘traverse’. They then returned to the corpus to check if these lemmas were used metaphorically. As a result, they were able to postulate four tentative source domains, among them building and motion. Next, they went back to their lists of a hundred collocates and collected further collocates suggesting these source domains, such as maison ‘house’ and progresser ‘progress’. They also compared their list of source domain lemmas against EuroWordNet which suggested connections and relationships between these and further lemmas. On the basis of these comparisons, they compiled lists of about 20 key lemmas conveying the source domains building and motion. Lastly, they searched for these lemmas in the corpus and found many new occurrences of metaphors.

Finally, Semino et al. (2005) exploited the semantic annotation tool in the online software Wmatrix software to compare the metaphors in a scientific journal, Nature Immunology, with a popular scientific journal, New Scientist (cf. Rayson 2015). Their software automatically categorized their data into semantic domains, and they then used it to see which semantic domains were over-represented in each journal. The idea was that some of these semantic domains could represent metaphorical source domains. They found out, for example, that the popular science articles were richer in some metaphoric source domains, such as war (e.g. ‘how the immune system’s front-line troops recognise the enemy’).

An example of current research: a case study on hope

Let us now turn to a case study of recent research. Building on an earlier diachronic study of metaphors of love in English (Tissari 2003) I set out to investigate metaphors of hope in four corpora – two representing Early Modern English and two representing Present-Day English: the 450,000-word Corpus of Early English Correspondence Sampler (CEECS, 1418–1680); the 551,000-word Early Modern English period of the Helsinki Corpus of English Texts (HCE, 1500–1710); the one-million-word corpora the Freiburg-LOB Corpus of British English (FLOB, 1991); and the Freiburg-Brown Corpus of American English (FROWN, 1991).

In an earlier unpublished study on Old English, Fabiszak and Hebda (2009) suggested that hope was a ‘peripheral emotion’: it shared some metaphors with other emotions, but not the emotion is force metaphor, because ‘hope is not conceptualised as a dynamic force influencing human lives at the present moment, but rather as a foundation for structuring these lives in the future’. I thus set out to see whether the same would apply to hope later in the history of English. My method was based on searching for particular target domain words in corpora. I chose to investigate the verb and noun hope and their derivatives to see what kind of metaphors of hope occurred with them.

To search for a word in historical English data is not as simple as to search for a word in Present-Day English data, because not only different forms but also many spelling variants occur. I therefore created an alphabetical word list of each historical corpus, read through the words beginning with ho- and collected the hope words. To be sure I covered all the spelling variants, I checked the Oxford English Dictionary (2015) for them. All the variant forms given for the noun and verb hope in it began with ho.
Eventually, the noun and verb *hope* and their derivatives were found to occur 603 times in CEECS, with many variants: *hoope*, *hoope*, *hop*, *hoped*, *hopefull*, *hopefullness*, *hopeing*, *hopeles*, *hopes*, *hopet*, *hopeth*, *hopid*, *hoping*, *hopinge*, *hoppe*, *hoppes*, *hopping*, *hops*, *hopyd*, *hopying* and *hopyt*. In all, the *hope* words occurred 1.34 times per 1,000 words in CEECS, which was far more often than in the other corpora. The number of occurrences of the *hope* words in the Early Modern English period of HCE was 222 and comprised the spelling variants: *hope*, *hoped*, *hopefull*, *hopeing*, *hopeles*, *hopes*, *hopeth*, *hoping*, *hopinge*, *hopynge* and *hoope*. This was 0.40 occurrences per 1,000 words. The total of *hope* words in FLOB was 341, which was 0.34 times per 1,000 words. Lastly, the total of *hope* words in FROWN was 320, yielding a frequency of 0.32 occurrences per 1,000 words.

Before any research on metaphor was conducted, there was an important finding: The *hope* words were more frequent in CEECS and HCE than in FLOB and FROWN, although the difference between HCE and the Present-Day English corpora was not as notable as that between CEECS and the rest of the corpora. A look at the data suggested that people often reported their own hopes in CEECS and HCE: 73 per cent of all the occurrences of the form *hope* in CEECS and 60 per cent of those in HCE occurred in the phrase *I hope*, while the corresponding figures for FLOB and FROWN did not rise over 30 per cent. It thus seemed that letters especially invited reports of personal hope, but also that the early corpus data as a whole behaved differently from the Present-Day English data. The same had already been observed in my research on *fear*, *love* and *shame* – these words were more frequent in CEECS and HCE than in FLOB and FROWN (Koivisto-Alanko and Tissari 2006: 196; Tissari 2006a: 144). Why this should be so remains something of a mystery: there might be one overarching reason, but the reasons could also vary from emotion to emotion.

I conducted the actual analysis by reading all the sentences in which the *hope* words occurred in order to see if hope was conceptualized metaphorically. Stefanowitsch (2006b) called this method metaphorical pattern analysis (MPA). He defines a metaphorical pattern as ‘a multi-word expression from a given source domain (SD) into which one or more specific lexical item [sic] from a given target domain (TD) have been inserted’ (Stefanowitsch 2006b: 66). For example, I identified a metaphorical pattern in the following sentence (the pattern is underlined, and the *hope* word in italics):

(4) Many others, though, have lost the houses that were home, that were built with hope, sweat, and large, scary mortgages.

*FROWN: The Miami Herald 19*

This pattern was considered metaphorical because houses are not literally built with hope, but, rather, with machines and tools. In line with an earlier study (Koivisto-Alanko and Tissari 2006: 197–8), I labelled the relevant source domain INSTRUMENT and the underlying conceptual metaphor HOPE IS AN INSTRUMENT.

The naming of the metaphors was nevertheless a challenging task. A choice to call a particular source domain INSTRUMENT, for example, ruled out other options. Example (4) could also have been analysed as HOPE IS A TOOL or HOPE IS A MACHINE. Even HOPE IS BUILDING MATERIAL and HOPE IS MONEY appear viable. I googled *build a house with* and found websites dedicated to building houses with hemp, pallets, straw etc., while a search for *built with* in the BNC produced examples where, for example, schools were ‘built with public rather than private money’ (*Converting Old Buildings* 1985–1993: 772). For the sake of comparison, example (5) is an Early Modern instance of the metaphor HOPE IS AN INSTRUMENT:
(5) Sir, I was muche comforted and refreshed with hope that, by your good meanes and your sonne’s, my brother should have had present delyverye from his long and tedious imprisonment . . .

CEECS: 1627 Talbot Bowes 315

It should be mentioned that the metaphorical patterns I identified were not always multi-word expressions. Even Stefanowitsch (2006b: 74) accepted burning anger and fierce anger as metaphorical, although they attested only one source domain word. A relevant example from the present data was:

(6) A solid performance, particularly up front, will give Wales hope and open the way for a better World Cup challenge next month.

FLOB: Evening Standard 23

This metaphor, signalled by the verb give, was labelled hope is a valuable commodity. It was modelled after the metaphor love is a valuable commodity (cf. Kövecses 1986: 95; Tissari 2003: 159, 380).

Eight main categories of metaphors were distinguished in the data: hope is a valuable commodity, hope is a container, hope is a fluid in a container (in the body), hope is an instrument, hope is a measurable quantity, hope is an inanimate entity (other than e.g. commodity, container), hope is an animate entity and hope is up. I have already given examples of two metaphors. Below, I will give examples of the rest.

The metaphor hope is a container suggested that someone can be inside hope:

(7) But we lyue in hope that we shall shortly receiue you againe, I pray God hartely we may, if it be his holy wyll.

HCE: Private letter by Margaret Roper 511

The metaphor hope is a fluid in a container (the body) occurred only once in the early data, but several times in the Present-Day English data:

(8) For poor people, seeking medical care from the marketplace drains them of hope and resources (Trevino and Moss 1983).

FROWN: See, ‘African-American Society and Education’ 22

The metaphor hope is a measurable quantity was frequent in the early data, but rare in the Present-Day English data:

(9) As for ther beeginninge, it is but vntoward, wth lyttle hope of good suckses.

CEECS: 1625 Thomas Meautys 1 117

(10) Still, reading autobiograpy, while it does not increase one’s hope for the race, does lend vast amusement in watching it all pass in review from the rail.

FROWN: Epstein, ‘First Person Singular’ 15

I had not used the source domain label measurable quantity in my previous research, but I had paid attention to emotions being quantified (Tissari 2003: 336–8; Tissari 2006a: 147, 149; Tissari 2006b: 24, 26, 33, 41). In a way, this metaphor went together with hope is a valuable commodity, because the more hope there was, the better. It was nevertheless important to distinguish between these two metaphors, because greater amount does not
always equal greater value. Consider shame, for example: More shame tends not to be better than less shame.

The metaphor category hope is an inanimate entity (other than e.g. commodity, container) comprised sub-metaphors such as hope is a rock and hope is a fragile object:

(11) . . . God doeth nothing else but lead vs along by the hand, til he haue setled vs vpon the rocke of an assured hope, that no one iote or title of his word shall passe till all be fulfilled?

HCE: Hooker, Two sermons upon part of S. Judes Epistle 8

(12) But councillors in Warwick have dashed the club’s hopes.

FLOB: Coventry Evening Telegraph 7

The metaphor label hope is an animate being did not specify whether this entity was a human, an animal or a plant. I had made such distinctions in previous research but found them problematic (e.g. Tissari 2003: 372–3). In example (13), hope was clearly an animal, but in example (18), it could be any animate being:

(13) Hope is a curtall-dog in some affaires: Sir Iohn affects thy wife.

HCE: Shakespeare: The Merry Wives of Windsor 44

(14) Faith and charity were in eclipse and even hope died at last.

FLOB: Candour 8

The metaphor hope is up is related to the metaphors happy is up and good is up which, according to Lakoff and Johnson (1980: 14–21) are based on physiological well-being being signalled by an upward posture:

(15) True, Linda had had a few boyfriends and had high hopes of at least two of them resulting in the kind of close, loving, long-term relationship she really wanted, but somehow it didn’t happen.

FLOB: Dryden, How to Untangle Your Emotional Knots 23

Table 8.1 shows the occurrences of each metaphor category per corpus. These are reported as instances per 10,000 words in each corpus, because the Late Middle and Early Modern English corpora are not directly comparable with the Present-Day English corpora, being only about half the size of FLOB and FROWN.

<table>
<thead>
<tr>
<th>Corpus</th>
<th>COMM</th>
<th>CONT</th>
<th>FLUID</th>
<th>INSTR</th>
<th>QUANT</th>
<th>INAN</th>
<th>ANIM</th>
<th>UP</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCE</td>
<td>0.22</td>
<td>0.27</td>
<td>0.00</td>
<td>0.05</td>
<td>0.40</td>
<td>0.15</td>
<td>0.07</td>
<td>0.02</td>
<td>1.18</td>
</tr>
<tr>
<td>CEECS</td>
<td>0.73</td>
<td>0.87</td>
<td>0.02</td>
<td>0.20</td>
<td>0.93</td>
<td>0.13</td>
<td>0.11</td>
<td>0.07</td>
<td>3.07</td>
</tr>
<tr>
<td>Total HCE + CEECS</td>
<td>0.45</td>
<td>0.54</td>
<td>0.01</td>
<td>0.12</td>
<td>0.64</td>
<td>0.14</td>
<td>0.09</td>
<td>0.04</td>
<td>2.03</td>
</tr>
<tr>
<td>FLOB</td>
<td>0.21</td>
<td>0.13</td>
<td>0.04</td>
<td>0.04</td>
<td>0.02</td>
<td>0.15</td>
<td>0.04</td>
<td>0.05</td>
<td>0.67</td>
</tr>
<tr>
<td>FROWN</td>
<td>0.21</td>
<td>0.16</td>
<td>0.02</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.06</td>
<td>0.03</td>
<td>0.60</td>
</tr>
<tr>
<td>Total FLOB + FROWN</td>
<td>0.21</td>
<td>0.15</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.10</td>
<td>0.05</td>
<td>0.04</td>
<td>0.64</td>
</tr>
</tbody>
</table>

Abbreviations of source domains: comm = valuable commodity, cont = container, fluid = fluid in a container, instr = instrument, quant = measurable quantity, inan = inanimate entity (other), anim = animate entity and up = up.
I used raw figures of metaphors per corpus to statistically compare the findings for the two periods with each other. A chi square test suggested that there was a statistically highly significant difference between them ($x^2 = 47.65857$, df = 7, $p = 4.15E-08$). In other words, the meaning of the *hope* words, “measured” in terms of metaphors, had changed between the two periods studied.

Table 8.1 can be used to zoom in on the differences between the two periods. It seems that these differences were greater with respect to some metaphor categories than others. There were more than twice as many *valuable commodity* metaphors, three times as many *instrument* metaphors, more than three and a half times as many *container* metaphors and 21 times as many *measurable quantity* metaphors in the early data than in the Present-Day English data. The only metaphor that was more frequent in the Present-Day English data than in the early data was *hope is a fluid in a container*, which was three times more frequent in the former than in the latter.

Table 8.1 also suggests that the difference between HCE and the Present-Day English corpora was small as regards the metaphors *hope is a valuable commodity* and *hope is an instrument*, but bigger as regards the metaphor *hope is a container*, and substantial as regards the metaphor *hope is a measurable quantity*. This indicated that the meaning of the *hope* words may have changed less than a comparison between CEECS and the Present-Day English data would suggest.

Taking the data as a whole, to what extent is *hope* metaphorically conceptualized like other emotions? Lakoff, Espenson and Schwartz’s Master Metaphor List (1991: 140–7) contains nine general metaphors for emotion, such as *emotions are forces* (*She was carried away by the song*) and *intense emotions are heat* (*The crowd was all fired up*). Of these, only three occur in my data as a whole: *emotions are entities within a person* (*hope is a fluid in a container*), *emotions are locations* (*hope is a container*) and *emotional self is a brittle object* (*≈hope is a fragile object*), the *hope is a container* metaphor being a subcategory of the more general *states are containers* metaphor (Lakoff and Johnson 1980: 30). The Master Metaphor List also listed three metaphors of hope: *hope is a beneficial possession* (*I have hope that he will return*), *hope is a child* (*I’m nursing a hope for a better life*) and *hope is light* (*He has bright hopes*; Lakoff, Espenson and Schwartz 1991: 151–2). Of these, only the metaphor *hope is a beneficial possession* (*hope is a valuable commodity*) occurred in the present data.

The present data corroborated Fabiszak and Hebda’s suggestion (2009) that hope was a peripheral emotion which did not attest all the metaphors of emotions, and particularly not the *emotions are forces* metaphor. This applied to all the data. However, as usual in research, this study also suggested new research questions, especially the following: (1) It would have been possible to compare the metaphors found in the present data with other listings of metaphors of emotions, for example Kövecses’s description of emotion (1990: 144–81). That being the case, how do we know which list is closest to describing emotions as against other concepts? (2) If *hope* is not conceptualized like an emotion, is it conceptualized like something else, for example, like a virtue? Question two also has a follow-up question which has not been researched within cognitive linguistics: How is virtue conceptualized?

A further question is whether we would have found all the metaphors characterizing emotions in the vicinity of *hope* words if we had had large enough data. As Stefanowitsch (2006b: 91) suggested: ‘[G]iven a large enough corpus, all metaphors will be instantiated for all emotions’. I used the examples in the Master Metaphor List to Google words and phrases (underlined) occurring with the noun *hope*, and found all the missing metaphors for emotions online: *emotions are entities within a person* (*God wants you to be filled with hope*), *emotions are forces* (*Carried Forward By Hope, a book title*), *intense emotions are heat* (*Burning Hope Ministries*), *emotion is motion* (*Moved by Hope: Burlington Hope Run/Walk set for May 17*), *effect on emotional self is contact with physical self* (*Touched By Hope, free medical clinic*),
Corpus linguistics and metaphor

Strong emotions are madness (*Dreams Gone Mad With Hope*, book title⁹) and strong emotion is blinding (*Blinded by Hope, Dazzled by Detail*, title of a CNN news story¹). This confirmed Stefanowitsch’s suggestion (2006b: 91). However, it seemed that several of these metaphors were what you might call creative metaphors. The metaphors which were gleaned from the corpora were thus likely to represent the more frequent varieties of metaphors of hope.

Reflections on the past, present and future of corpus linguistic research on metaphors

When I was a PhD student at the turn of the millennium, one of my colleagues said that I could not use electronic corpora to study metaphors. Today, however, it is no longer a novel idea to use an electronic corpus to study metaphors. To mention a milestone in the field, Stefanowitsch and Gries (2006) were able to bring together a number of researchers combining research on metaphor with corpus linguistics. Their volume combined softer, more qualitative approaches to the topic such as Semino’s work on metaphors for speech activity (2006) with more quantitative and computational work such as Martin’s on context effects on metaphor comprehension (2006). More importantly, Stefanowitsch and Gries (2006) were able to show that corpus linguistics had become part and parcel of metaphor research and that there were many potential ways of approaching metaphors in corpora, ranging from considering metaphors as fuzzy categories (Hanks 2006) to including the study of metaphor in discourse analysis (Koller 2006; Partington 2006).

Moreover, Stefanowitsch (2006a, b) was able to suggest that there was still much to do in the field. For example, he proposed that the method of searching for metaphors occurring together with target domain lexis should be systematically applied to a large number of concepts (Stefanowitsch 2006b: 102–3). He also envisioned that corpus-based research on metaphor should involve more elaborate statistical methods (Stefanowitsch 2006a: 12). These goals still seem worth aspiring to. The first one relates to my case study on hope: we still do not have exhaustible corpus evidence of which metaphors distinguish emotions from other concepts, for example. The second one relates to what has been called the “quantitative turn” in cognitive linguistics (Janda 2013), which is increasingly affecting metaphor research.

What I have to say about the future of corpus linguistic research on metaphor concerns small corpora and big corpora or, even more generally, small data and big data. We have just seen how even small corpora can be used to study conceptual metaphors and how metaphor can be an indicator of semantic change. We indeed seem to be at an interesting juncture where things are happening both on a small scale and a large scale. Small data can be analysed by hand when there is a need to conduct analyses which computers cannot perform yet. In particular, new approaches to metaphors can be tested on small data. Examples of important “experiments” on small data include the Hamburg Metaphor Database whose compilers collected and analysed metaphors from a French corpus to create a metaphor resource which could be used, for instance, to predict metaphors in other data (Lönneker-Rodman 2008); and the even more recent VU Amsterdam Metaphor Corpus which is the ‘largest available corpus hand-annotated for all metaphorical language use, regardless of lexical field or semantic domain’, ‘based on a systematic and explicit metaphor identification protocol’ and which ‘covers about 190,000 lexical units from a subset of four broad registers from the BNC-Baby’ (Welcome to the VU Amsterdam Metaphor Corpus Online! 2015; see also Steen et al. 2010).

Even small data do not exclude statistical analysis, but it is big data which especially require quantitative methods. As bigger and bigger data become available, research on metaphor is likely to develop towards more and more linguistic computing. And vice versa, advances in linguistic computing facilitate the conducting of research on metaphor on a large
scale (e.g. see the ‘Mapping Metaphor’ project described in Chapter 16). It is to be expected that data mining techniques will be used more and more to find metaphor in big data and that the automatic detection of metaphor will be significantly improved even in the near future. To conclude, if you want to be innovative in corpus-based metaphor research, I suggest that you study a concept that has not been studied yet, do something that has not been manually done yet, or venture to develop the large-scale statistical analysis of metaphors.

Notes

1 From now on, I will use the term corpus to refer to an electronic corpus.
3 I thank prof. Dirk Geeraerts for this suggestion.
4 On 23 June 2015.
6 http://www.amazon.com/Carried-Forward-Bregdan-Chronicles-Historical-ebook/dp/B00K6IBRYE.

Further reading


References

British National Corpus (BNC), The. Online. Available at: http://corpus.byu.edu/bnc/.


