



Alexander, M., and Bramwell, E. (2014) Mapping metaphors of wealth and want: a digital approach. In: Mills, M., Pidd, M. and Ward, E.(eds.) *Proceedings of the Digital Humanities Congress 2012*. Series: Studies in the digital humanities. The University of Sheffield.

Copyright © 2014 The Authors

<http://eprints.gla.ac.uk/89963/>

Deposited on: 09 February 2015

Enlighten – Research publications by members of the University of Glasgow
<http://eprints.gla.ac.uk>

Mapping Metaphors of Wealth and Want: A Digital Approach

by Marc Alexander and Ellen Bramwell

Citation

Alexander, Marc and Ellen Bramwell. 'Mapping Metaphors of Wealth and Want: A Digital Approach'. In: Clare Mills, Michael Pidd and Esther Ward. *Proceedings of the Digital Humanities Congress 2012*. Studies in the Digital Humanities. Sheffield: HRI Online Publications, 2014. Available online at: <<http://www.hrionline.ac.uk/openbook/chapter/dhc2012-alexander>>

Abstract

The AHRC-funded *Mapping Metaphor with the Historical Thesaurus* project aims to provide data on the extent and development of metaphor across the history of English. It uses the full database of the *Historical Thesaurus of English*, which extensively categorises and classifies the recorded vocabulary of the English language from Old English to the present day. By using this database to map semantic categories onto one another, and thus showing lexical overlap in different conceptual fields, we aim in the project to provide results which will demonstrate the widespread, systematic and far-reaching impact of metaphor on English.

This paper outlines the digital and linguistic methodologies used by the project, and presents a case study of the semantic categories of *wealth* and *poverty*, demonstrating the metaphorical links between these categories and the rest of the language. In addition, we discuss the nature of lexical overlap as we use it in the project, and discuss both the quantitative and diachronic dimensions of the data we are manipulating and their implications for projects of this type.

Mapping Metaphors of Wealth and Want: A Digital Approach

by Marc Alexander and Ellen Bramwell

1. Introduction

"...unless you are at home in the metaphor, unless you have had your proper poetical education in the metaphor, you are not safe anywhere. Because you are not at ease with figurative values: you don't know the metaphor in its strength and its weakness. You don't know how far you may expect to ride it and when it may break down with you. You are not safe with science; you are not safe in history." (Frost 106)

The poet Robert Frost, writing in 1931, emphasised the importance of metaphor not simply as a rhetorical device but rather as the "whole of thinking" (*ibid* 104). This point, echoed in linguistics throughout the twentieth and twenty-first centuries as the importance of cognitive science became apparent in the analysis of language, is particularly relevant to the digital humanities in the present day. As Frost notes, much of language is figurative, and in fact more recent research has shown somewhere between 8% and 18% of English discourse is metaphorical, with an average of every seventh word being a metaphor (Steen et al 765). To apply this to a dataset such as the *British National Corpus* (100 million words, 1980s-1993), we can expect to find in that corpus 14.3 million words used figuratively, including some very common expressions as time is space (That's all *behind* us now, *Next* week and the week *following* it), having control is up (I am *on top* of the situation, He's *at the height* of his power, He is *under* my control, He is my social *inferior*), theories are buildings (The theory needs more *support*, We need to *construct* a *strong* argument for that, The theory will *stand* or *fall* on the *strength* of that argument, We will show that theory to be without *foundation*), and understanding is seeing (We will now *show* that... We ought to *point out* that... I don't *see* your point *in* that argument, Your argument is just *not clear* enough; all the preceding examples are from Lakoff and Johnson 1980).

This fact is a concern, given that our ability to handle non-literal language in digital humanities is not yet fully formed. While advances are being made in the semantics of digital texts, alongside emerging concepts of a semantically-aware Web, we are at a very early stage in comprehensively and systematically understanding English metaphor, and therefore at an early stage of being able to accurately deal digitally with the meanings encoded in those texts. The current leading semantic tagger of English, USAS, currently deals with metaphor not as links between meanings (properly, links between *semantic domains*), but rather treats metaphorical terms as polysemous words, so that a metaphorical word appears twice in its database, once with its literal meaning and once with its figurative meaning (see further Archer, Wilson, & Rayson). The present article describes a novel methodology for systematically identifying metaphors in English using a digital data-driven approach in conjunction with manual tagging and filtering, and points the way towards a complete database of English metaphor. The procedure here described was piloted in Alexander and Kay 2011 [2010], and forms the core of the AHRC-funded *Mapping Metaphor with the Historical Thesaurus* project at the University of Glasgow. The following subsections describe some aspects of this project and the data it uses, while the remainder of the article focuses on metaphors for *wealth* and *poverty* found in the dataset, and the implications and issues this approach engenders for a digital approach to meaning.

1.1. The Dataset – The Historical Thesaurus of English

The project takes as its core data the *Historical Thesaurus of English*, published in 2009 as the *Historical Thesaurus of the Oxford English Dictionary*. The *Historical Thesaurus* (hereafter HT) is the world's only historical thesaurus, covering English from Anglo-Saxon times to the present day, and is the largest semantic database ever constructed for any language. It contains over 790,000 word entries, each representing a particular word sense, within 240,000 categories which cluster these words under the distinct concepts they refer to. This dataset is larger than the comprehensive *Oxford English Dictionary* (Simpson and Weiner), which contains 616,500 senses (Algeo 137). One further uniqueness is the fine-grained hierarchical layout of the data categories; each conceptual set of words is nested within other, wider categories, so that, for example, the verb category *Live dissolutely* is within *Licentiousness*, itself adjacent to *Guilt and Rascalry* and within the wider category *Morality*. Each individual point in the hierarchy can contain both word entries for the concept represented by that point and also all the conceptual descendants which follow it, each surrounded by

siblings of similar meaning.

The HT dataset therefore contains all the meanings recorded in the history of English and also contains all the words we know to have been used to instantiate these meanings. This is particularly important for the metaphor identification methodology used in the Mapping Metaphor project, which treats word overlap between these semantic categories as an indicator of likely metaphoricality for further analysis, and this process is described in more detail below.

1.2. The Project - Mapping Metaphor with the Historical Thesaurus

The Mapping Metaphor project has been designed to harness the power of the HT for research into patterns within the English language over time and semantic space. The idea of patterns is crucial here, with the aim of the project to discover (or, metaphorically, to 'map') those domains of meaning which are used to talk about other domains of meaning. As a result of the comprehensive nature of the HT, which provides us with a historical-semantic record of the English language in its near-entirety, this task can now be completed in a systematic manner, identifying all links for which evidence for metaphor is present in the source material.

The specific process for identifying metaphor in the lexical data is discussed in more detail in section 3 below, but the basis of the project involves dividing the HT into Mapping Metaphor categories, each representing a discrete domain of meaning. These domains are based on the HT classification, but are different in that they overlay the HT categories in order to provide a broad and manageable set. At its finest level of detail, the HT has almost a quarter of a million categories; the Mapping Metaphor team have split these into 411 separate domains, each largely based on higher-level categories in the HT hierarchy. For example, the Mapping Metaphor category *N02 Wealth* contains all the HT data from category reference 02.07.05 (The Social World (02) > Possession (07) > Wealth (05)) to the end of 02.07.05.01 (The Social World (02) > Possession (07) > Wealth (05) > Riches (01)) in the HT classification. This consists of 509 separate lexical items. When the HT hierarchy reaches 02.07.06, where the first category heading is 'Poverty', the Mapping Metaphor team have begun a new category, *N03 Poverty*. This changes the nature of the data from strictly hierarchical to discrete 'chunks' which, crucially for our purposes, are broadly comparable in size and breadth of coverage.

The aim of the project, as briefly touched on above, is to find which of these

separate domains are linked in metaphorical relationships, and the comprehensive and unique nature of our data enables us to make a systematic and methodical investigation with the results entirely underpinned by lexical evidence. The results will include a publicly-available website, which will allow others to explore these metaphorical links across the history of English.

The further implications for the digital humanities of this project are, we hope, evident. Many studies in DH and in corpus linguistics focus on lexical data, and without a clear understanding of metaphor in this data, and an ability to manipulate it, we are working with an incomplete toolset. The discussion below of the issues metaphor raises for the digital humanities should serve as an illustration and worked example of this problem, and a demonstration of some solutions the project has arrived at for the problem of working with lexical data.

2. *Wealth and Poverty: An Overview*

The following two tables show a selected summary of the data obtained using this procedure for the Mapping Metaphor categories *N02 Wealth* and *N03 Poverty*. Each table is sorted by the Mapping Metaphor category which has a lexical overlap with the 'home' category for each table. These tables summarise the raw data found in the original files created for analysis (samples of some of this raw data are also found in Table 3 and Table 4 below).

Table 1: Summary of *N02 Wealth's* overlap with the remainder of the HT database.

Overlap Category	Example Words	Notes
<i>A07 Wild/uncultivated land</i>	rich, richness, fat, strong, wanton	Refers mainly to fertile land.
<i>A13 Flow/flowing</i>	affluent (as flowing), increase (water level), confluent	Actions and states of water accumulation are similar to those of monetary accumulation.
<i>B06 Health and disease</i>	well, strong, solid	All relating to good health.

<i>B28 Bodily shape/physique</i>	fat, plum, pousy, full, opulent, fatten; cob (= fat person)	In modern times strange in contrast with <i>B06</i> , but in earlier times fat/healthy were not contradictory when contrasted with impoverished.
<i>B73 Food</i>	Snug	Comfort-related.
<i>F29 Sufficient quantity</i>		Not a metaphor (hyperosemy)
<i>H27 Attention, judgement</i>	solid (of people/judgments and US slang); juicy, plenty, solid); enrich(ing/ment)	Judgment of worth; mainly refers to things/people which are good, excellent, worthy, acceptable. Includes a nice run of early C20th slang for <i>excellent</i> .
<i>Y09 Money</i>		Not a metaphor (metonymy).

N02 Wealth has a clear relationship with largeness, increase, and bounty, such as that which can be found in landscapes (*A07*), people (*B28*), and health (*B06*). The connection between wealth and a large bodily shape (*fat, full, cob*) is an embodied metaphor, where there is a clear sensation link between the positive feeling of being wealthy and that of being pleasantly full of food, as well as a causal link between persons with wealth and persons who can afford to eat well. This is a reflection of the diachronic data under analysis – just as *cob* is no longer common in modern English, there is a shift in attitudes regarding the stereotype of a wealthy person. Two interesting items here are not metaphorical and are discussed below – *F29 Sufficient quantity*, here marked as representing the phenomenon of hyperosemy, which we discuss below, and *Y09 Money*, which is an example of metonymy.

Table 2: Summary of *N03 Poverty's* overlap with the remainder of the HT database.

Overlap Category	Example Words	Notes
<i>B07 Ill-health</i>	weak, poorly, decay, waste	

<i>B28 Bodily shape/physique</i>	pinched, starved, withered, poorness, feeble	
<i>D38 Matter, bad condition of</i>	waste, decay	
<i>E03 Destruction</i>	ruin, waste	
<i>E23 Harm/injury/detriment</i>	mischief	This is an early sense, found in Middle English - as an example, the <i>OED</i> cites Henry VI's 1433 Rolls of Parliament (424/2) "They bee nowe in grete myschief and necessite".
<i>E24 Adversity/affliction</i>		Not a metaphor (hyperosemy).
<i>E25 Failure/lack of success</i>	default, want, mischief	
<i>E45 Position, relative</i>	bare, stark, skinned	The category title obscures the link here between nakedness and poverty.
<i>H31 Contempt</i>	beggar, pinch, cheapo, bankrupt, lowness, ruin, poorly	
<i>I06 Mental pain/suffering</i>	stony, miserable	'Stony' is an interesting one-off metaphor, as in being petrified with grief.
<i>I15 Humility</i>	lowness, embarrassed, broken, poorly	
<i>O03 Speech/act of speaking</i>	beg	Not a metaphor (metonymy).
<i>T05 Moral evil</i>	naught, ruin, fall, mean	

<i>V03 Church government</i>	poor friars	Example of noise.
------------------------------	-------------	-------------------

As an inverse of *N02 Wealth*, *N03 Poverty* naturally has clear associations with absence, scarcity and paucity. Interestingly, though, it shows clear evidence here of a lack of money being tied to a lack of other important qualities, such as success (E25), esteem (H31), happiness (I06), pride (I15), and morality (T05). The social implications for the clear bias in how English speakers across the past millennium have conceptualised and analogised the notion of poverty are significant.¹

These two tables suffice as samples of the data which arises from a comparison of lexical overlap in a database such as the HT. The following sections discuss the implications which the interpretation of this data has for DH.

3. Issues Arising

As the *Mapping Metaphor* project engages simultaneously with the history of English, the stylistic dimension of metaphor and the digital analysis of lexical data, the project has had to confront a number of significant theoretical and methodological issues surrounding metaphoricity in English. In this section, we will discuss those issues which arise most particularly in the digital humanities, and illustrate the conclusions which data such as that outlined in Section 2 have led us to.

3.1. What can Lexical Overlap be?

We here discuss three possibilities for what computationally-identified lexical overlap can be, beginning with the metaphors which the project itself is most concerned with, but also discussing further categories which researchers have to differentiate from metaphor.

3.2. Lexical Overlap as Metaphor

Analysing lexical overlap, in the context of this project, consists of taking every word in the start category (such as *N02 Wealth* or *N03 Poverty*) mapped against the entirety of the remainder of the HT (that is, everything which is not the start category). The results from this query consist, therefore, of lists of words shared between the start category and any other. Each word is represented individually in the database results, along with its

individual meaning from its non-start category.

While this may sound like an ideal 'big data' project, with easily obtainable results gained computationally, the actual process involved in gaining useful results is not as straightforward. As with so much semantic and lexical research, this overlap data is only the very first stage in a much more complex analysis.

Section 2 above described the results of the second and third stages of qualitative analysis, where the relationships between categories have been coded and a wider picture is beginning to emerge of the large-scale connections within the data. To get to this stage, the first stages of qualitative analysis consist of the computationally-obtained data being examined and coded manually by the project team. This section will use the metaphorical overlap between *N02 Wealth* and *N03 Poverty* and *B28 Bodily shape/physique*, from above, as an example, and will discuss some of the theoretical issues present in this analysis.

The data at this first stage consists of a list of category names with various statistics pertaining to each category, as discussed in Section 3.2 below. This spreadsheet is used to label whole categories as being metaphorical or otherwise. What these decisions are based on though is the lexical overlap corresponding to each of these category names. The full overlap data from *N02 Wealth* and *N03 Poverty* to *B28 Bodily shape/physique* are reproduced in tables 3 and 4, below.²

Table 3: Lexical overlap between *N02 Wealth* and *B28 Bodily shape/physique*.

B28	solid	Aj	1741-	.robust
B28	fat	Vr	1567 also fig.	.fat/plump
B28	cob	N	1583	..person
B28	stock	Vi	1808 Scots	Loose/stiff condition
B28	strong	Aj	OE-	Physically strong
B28	big	Aj	a1300-1599	Physically strong
B28	strong	Aj	a1225-	.robust
B28	opulent	Aj	1896	.fat/plump
B28	strong	Aj	1398-	..of vital organs/functions

B28	strong	Aj	1398-	.characterized by use of strength
B28	plum	Vt	1594	.fat/plump
B28	fatten	Vi	OE + 1676- also fig.	.fat/plump
B28	full	Aj	1577-	.fat/plump
B28	pursy	Aj	1576- also fig.	.fat/plump
B28	plum	Aj	1570-1594	.fat/plump
B28	fat	Aj	OE-	.fat/plump
B28	fat	N	1726-	..state of having
B28	full	Aj	1577-	.rounded
B28	make	N	1719-	Bodily shape/physique
B28	fat	Vi	a1225-1825	.fat/plump

Table 4: Lexical overlap between N03 Poverty and B28 Bodily shape/physique.

B28	feeble	n	1340 + 1833-1896	.one who is weak
B28	stump	n	1875	..person
B28	starkness	n	c1440-	Loose/stiff condition
B28	stiff	aj	c1305-	Loose/stiff condition
B28	sturdy	n	1895	..person
B28	stiff	aj	1297-a1677	.robust
B28	sturdy	aj	c1386-	.sturdy
B28	feeble	aj	c1175-	Physically weak
B28	lowness	n	1638	Shortness
B28	feeble	vt	a1340-1614	Weaken
B28	stiff	vi	1399	Become strong (of the body/its parts)
B28	poorness	n	1577	..state of having
B28	extenuate	vt	1533-1887	.thin
B28	limit	n	1636(2)	Bodily shape/physique

B28	sturdy	aj	c1386-	Broad
B28	leanness	n	OE-	..state of having
B28	poverty	n	1523-	..state of having
B28	starveling	n	1546- also transf. & fig.	..person having
B28	extenuate	aj	1528 + 1689	.thin
B28	starved	aj	1597-	.thin
B28	pinched	aj	1614-	.thin
B28	withered	aj	a1500/34-	.shrunk
B28	waste	vi	1763-	..by training
B28	feeble	vi	a1225-1496 + 1889 arch.	Become weak

The researcher deciding on a code for each category is interested in whether they can find evidence for a metaphorical connection (or transfer) between the two categories. These lexical items provide the sole evidence for or against this, and so the method is, above all, data-driven.

However, this does rather raise the question of what a metaphor is and how one might be recognised. For the purposes of the initial data-analysis stage, a lexical item is metaphorical where the vocabulary from one category is being used in another category in a non-literal way - that is to say, where attributes of one concept are being transferred onto another. (This does not include the related concept of metonymy, however, discussed below in section 3.3.2, which tends to be a lexical phenomenon.) This brings us back to the aim of the project: to find conceptual links between semantic domains which show evidence of systematic metaphor. We need a critical mass of words for categories to show conceptual metaphor, not just single words. However, as lexical items are our evidence, the project developed the additional option of coding categories which contain few instances of metaphor as being 'weak' links, therefore giving a fuller record.

One further issue for a data-driven approach is that in many respects it is easier to ostensibly recognise metaphor than to clearly explain why it is metaphorical. In the raw data above, and as shown in section 2, the links between *B28 Bodily shape/physique* and both *N02 Wealth* and *N03 Poverty* are strongly metaphorical. The individual lexical links here are plentiful, and certainly form enough of an evidence base to describe them as strong

conceptual links. They are also nicely symmetrical, linking with different aspects of the body shape category to form polar opposites of wealth and poverty, strong and fat, and weak and thin. The direction of metaphor is also important to our later analysis of the links, and it is clear from this evidence that words from the domain of body shape are generally being used to talk about wealth and poverty, rather than the other way round.³ However, there are also examples in the other direction, with bodily shape including *opulent* and *poverty*, with the bidirectionality illustrating the strength of the conceptual links here.

The value of this method is that it does not look only for metaphoricity in lexical items, which might be useful at the level of individual texts. Rather, it builds this lexical data into an evidence base for systematic metaphorical links between *categories*, and does this in a way which is itself systematic across English. In this way, with database queries providing data for every possible combination of categories (a theoretical maximum of 320,000 category pairs, although the actual number is luckily lower than this), the technique described here analyses metaphor systematically and methodically to produce, scheduled for late 2014, a full dataset describing every metaphorical link captured in the lexical record of English over the past thirteen centuries.

3.3. Lexical Overlap as Other Semantic Phenomena

This core aim aside, the Mapping Metaphor procedure for dealing with lexical overlap which is not metaphorical but nonetheless has evidence of a clear semantic link between the categories is for this overlap to be coded as *relevant* to that domain. For the purposes of this article, and future research, these relevant links can generally be split into two major categories of interest, outlined below.

3.3.1. Hyperosemy

We here propose the term *hyperosemy*, as first used in Alexander 2011, as a way of referring to the ways in which the HT includes categories which are generic antecedents of many other categories. For example, the category of *N02 Wealth* shares a number of words with *F29 Sufficient quantity*, and yet this link is neither metaphorical nor to be discarded. Instead, F29 is a more generic instantiation of the same concept (wealth is sufficient quantity of money). A single hierarchy cannot possibly allow wealth to be a descendant of possession and a descendant of money and a descendant of quantity simultaneously, and so a concept's antecedents are found in multiple places.

We use the term hyperosemy (a semantic equivalent of hyperonymy, used of superordinate words such as *tree* in relation to *oak* and *maple*) to refer to this phenomenon. This neologism is a direct result of the digital approach to this dataset, which necessarily disrupts the established hierarchy of the HT dataset for its own purposes. We expect similar semantic terms to be needed in future digital research on the HT.

3.3.2. Metonymy

A more established term is metonymy, generally defined in semantics as a phenomenon similar to metaphor but one which refers with regards to contiguity rather than correlates with regards to similarity. What this generally means is that metonymy is the use of an element or attribute of something for that thing itself - for example, a person's name for their writings (*I read Shakespeare every day*), a part of someone's body for what it does (*Hold your tongue!* for *Stop speaking!*), a building for the people who work there (*The White House issued a statement yesterday*), and so forth. These highlight very proximate concepts, rather than connect distant ones like metaphor. *Y09 Money*, in this instance, has significant lexical overlap with *N02 Wealth*. The problem here for a digital approach to words is very tightly bound to the issue of multiple antecedents in the HT hierarchy as outlined above for hyperosemy; *Y09* is semantically proximate to *N02* but is not hierarchically proximate as it presently stands in the HT dataset. A huge benefit to the analysis of lexical overlap for metaphor studies, and one which is also potentially transformative to the digital humanities is the indexing of these non-metaphorical links between concepts alongside the metaphorical work.

3.4. Lexical Overlap as Noise

Finally, *noise* is what we here term non-metaphorically motivated overlap which is generally due to homonymy in the history of English. This problem is somewhat intractable; with a finite supply of consonants and vowels, combined with limiting phonotactic rules, there are only a finite number of possible English words. In many cases, identical word-forms arise through unrelated historical processes, leaving English with two unconnected words, with no etymological or conceptual link between them. Such words are of no interest to us in this project.⁴ An example is *breeze*, with the oldest word, meaning a gadfly, coming from Old English, a second, meaning a gentle wind (particularly northern or northwesterly) from Old Spanish in the fifteenth century, and the third word, meaning dust from burning bricks, likely from Old French somewhat later. These three distinct words share the same

word-form, meaning they appear to a computer to constitute lexical overlap but do not in fact have anything to do with each other.⁵ This then constitutes noise in our data. However, noise and metaphor often co-exist, and the second gentle-wind sense of *breeze* does have non-noise extensions (termed polysemy), such as the twentieth century slang term *breeze*, meaning something easy to achieve. This simultaneity means that the only way in which homonymy and metaphor can be distinguished in the HT dataset is through manual intervention. In future years, when the HT is fully linked with the OED and both are available for research as fully-accessible datasets, this problem will reduce substantially.

3.5. How can Quantitative Data Assist this Process?

In order to speed up the onerous and complex process of data analysis, and in common with most other big data projects, we experimented with using quantitative methods as a means of helping identify where metaphor might be more likely to occur, in order to focus coders' effort onto these areas.

From the lexical overlap data generated (Table 3 and Table 4), it is straightforward to generate simple descriptive statistics. Those used in the project included figures for the number of words in each category which overlapped with the start category; for the number of unique word forms this represents (rather than the same word repeated over and over; see, for example, *full* and *strong* repeatedly in Table 3); for the total number of lexical items in the Mapping Metaphor category being interrogated; and finally the earliest and latest citation dates of the overlap data. The statistics to do with size were useful in establishing how significant overlap really is in the context of its various categories; two very large categories will have many words in common, for example, but this overlap may contextually be a very small proportion of their overall size. The measure of unique words solved the problem where the size of overlap lists may be artificially inflated by very polysemous words with several semantically close senses - however, this measure conflated polysemous senses which might individually be counted as metaphors.

Overall, these sets of quantitative data allowed for experimentation as to whether different combinations of these measures might make metaphorical categories more visible. However, as can be seen from the discussion above, this was problematic and the results are frankly mixed. Calculations including all of these elements alerted us to categories which might be very relevant to the start category, or which might be more likely to have systematic metaphor if they have any at all. Importantly though, there was

no quantitative process which consistently identified highly metaphorical categories when trialed over the data from several start categories.

3.6. What Issues Arise from the Diachronic Dimension of the Data?

The HT dataset uses OED headwords for its entries – that is, the particular spelling and word form which is used by the OED itself (so while *ampoule*, a sealed vessel which contains sterile materials or medicine for injection, could also be spelled *ampul* or *ampule*, the OED has chosen *ampoule* as its headword, and the HT dataset will include it only under that form). However, the OED does not include Old English material which has no citation after 1150 (the beginning of the Middle English period), which is why the HT is significantly larger than the OED, supplemented as it is by other sources. Data from these other sources were published in 1995 by Kay and Roberts (as TOE), as a precursor to the 2009 HT, and as a consequence the HT contains two sets of headwords in its database – the main headwords under which a word is cited in the OED, and the TOE headwords for Old English material.

This causes an issue for studies of lexical overlap using the HT. Many words are given both an OE and an OED headword, so that the word book, for example, is recorded both as *book* and as *boc*. Some words have only an OE headword, because they did not survive past 1150 (such as *fremigendlic*, a synonym for advantageous, which is not found past the eleventh century). Many others only have an OED headword, as they are later than Old English. It is therefore necessary to capture four possible overlap types: one between two words only found in OE, one between two words only found post-OE (ie, using only OED forms), one between a word found only in OE and one found in both OE and later (so one word with only an OE form and another with both an OE and a non-OE form, and the metaphorical link exists only in OE) and one with overlap found in both its OE and its OED forms. This complex situation is best resolved by treating the OE data and the non-OE data as separate entities, analysing them separately (ie looking at OE to OE and then at non-OE to non-OE), and then linking these results together afterwards, merging data where necessary. This project is, as far as we know, one of the first to digitally use both Old English and later English data simultaneously to reach wide-scale conclusions about the language, and the separation of these two datasets for practical reasons is the first time we know of this having to occur.

4. Conclusion

The present article has described the issues which arise from a digital investigation of lexical overlap in a comprehensive semantic dataset, with the objective of using this overlap for the analysis of English metaphor. We have outlined some of the methodological issues and linguistic contributions originating from this research, in the knowledge that further alterations and development may yet be required to the full understanding of lexical overlap in these databases. Other studies, using similar but much smaller datasets, such as those on Roget's Thesaurus (Davidson) or WordNet, can also draw on the developments outlined here, and we keenly anticipate the completion of the full analysis of metaphor in the history of English by the full project team in coming years.

Finally, it is essential to emphasise that there is no way around the need, at this stage, for manual intervention and expert coding of metaphorical/non-metaphorical links. We strongly consider the Mapping Metaphor project to be a digital humanities project, but we recognise that while DH can provide huge amounts of new data for analysis, the intervention of a scholar is still essential in many areas - in the dominant paradigms of modern psycholinguistics (e.g. Pederson), meaning is strictly understood at its core as being idiolectal, and so ambiguity, even if not intentional, is not easily resolvable independent of an interpreting mind. While meaning studies is a challenging field, and one which we believe to be key to the future of the digital humanities, work such as that described in this article is still necessary to unlock the potential of semantic contributions to digitally-oriented research.

5. References

Alexander, Marc. *Research Implications of the Historical Thesaurus*. Conference paper at Historical Semantics, Etymology and Lexicography: A Meeting of the Philological Society, University of Glasgow, 2011.

Alexander, Marc & Christian Kay. *Mapping Metaphors Across Time with the Historical Thesaurus*. Conference paper at Helsinki Corpus Festival: The Past, Present, and Future of English Historical Corpora, University of Helsinki, Finland. Based on an earlier paper at The 3rd UK Cognitive Linguistics Conference, University of Hertfordshire, 2011 [2010].

Alexander, Marc & Andrew D. Struan. 'In countries so unciviliz'd as those?': The Language of Incivility and the British Experience of the World. In Martin Farr & Xavier Guégan (eds.) *Experiencing Imperialism: Interdisciplinary and transnational perspectives on the colonial and post-colonial British*. London: Palgrave Macmillan, 2013.

Algeo, John. *The Emperor's New Clothes: The second edition of the society's dictionary*. Transactions of the Philological Society, 1990. 88(2).

Archer, Dawn, Andrew Wilson, & Paul Rayson. *Introduction to the USAS Category System*. <http://ucrel.lancs.ac.uk/usas/usas%20guide.pdf>.

Davidson, George W. (ed.) 2002. *Roget's Thesaurus of English Words and Phrases: 150th anniversary edition*. London: Penguin, 2002.

Frost, Robert. *The Collected Prose of Robert Frost*, ed. by Mark Richardson. Harvard: Harvard University Press, 2007.

Kay, Christian, Jane Roberts, Michael Samuels, & Irene Wotherspoon (eds.). *Historical Thesaurus of the Oxford English Dictionary*. Oxford: Oxford University Press, 2009. Data available through <http://historicalthesaurus.arts.gla.ac.uk> and <http://www.oed.com>.

Lakoff, George & Mark Johnson. *Metaphors We Live By*. Chicago: University of Chicago Press, 1980.

Murray, James A. H., Henry Bradley, William A. Craigie, & Charles T. Onions (eds.). *A New English Dictionary on Historical Principles*. [=Oxford English Dictionary, 1st edn], 10 vols/128 fascicles. Oxford: Clarendon Press, 1884-1928.

Pederson, Eric. Cognitive Linguistics and Linguistic Relativity. In Dirk Geeraerts & Herbert Cuyckens (eds.) *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press, 2007.

Roberts, Jane & Christian Kay, with Lynne Grundy. *A Thesaurus of Old English*. Amsterdam: Rodopi, 2000 [1995].

Samuels, Michael. *Linguistic Evolution: With special reference to English*. Cambridge: Cambridge University Press, 1972.

Simpson, John A. & Edmund S. C. Weiner *et al* (eds.). *Oxford English Dictionary*. 2nd edn. Compiled by John A. Simpson & Edmund S. C. Weiner

from Murray *et al* 1884 and further supplements edited by William A. Craigie, Charles T. Onions, and Robert Burchfield, with additional corrections edited by John A. Simpson & Edmund S. C. Weiner. 20 vols. Oxford: Clarendon Press, 1989.

Steen, Gerard J., Aletta G. Dorst, J. Berenike Herrmann, Anna A. Kaal, & Tina Krennmayr. *Metaphor in Usage*. *Cognitive Linguistics*, 2010.21(4).

Footnotes

- ¹ A discussion of the social implications of the metaphorical construal of certain groups is beyond the scope of the present article. A discussion of these, with reference to HT data, can be found in Alexander and Struan 2013.
- ² This is an abridged version of the information available to the coding team, which includes each lexical item's location in the HT hierarchy.
- ³ The dates in these examples show the dates that these lexical items have been in use in *B28 Bodily shape/physique* and not the dates when they were metaphorically transferred to *N02 Wealth* and *N03 Poverty*, which will be shown when running the query in the other direction.
- ⁴ The once-pervasive view that homonymy is of no interest to linguists other than as an oddity is challenged in Samuels 1972: 67-77, who demonstrates the use of homonymic clash as a regulatory mechanism in the evolution of a language.
- ⁵ Linguists often deal with this problem by appending subscript labels, one for each distinct homonym family (viz. breeze₁, breeze₂, and breeze₃). We do not use this mechanism in this project for technical reasons.