

WORD FREQUENCY AND KEYWORD EXTRACTION

AHRC ICT Methods Network Expert Seminar on Linguistics
8 September 2006, Lancaster University, UK

Love – ‘a familiar or a devil’? An Exploration of Key Domains in Shakespeare’s Comedies and Tragedies

Dawn Archer, *University of Central Lancashire, UK*; Jonathan Culpeper, *Lancaster University, UK*;
Paul Rayson, *Lancaster University, UK*.

Keywords

love, Shakespeare, *love-tragedies*, *love-comedies*, conceptual metaphors, keywords, key domains, corpus, semantic annotation, collocates, quantitative, qualitative, aboutness, content analysis.

Abstract

Researchers increasingly use corpus linguistic methodologies such as keyword analysis to study Shakespeare (see, for example, Culpeper 2002) or are studying Shakespeare from the perspective of cognitive metaphor theory (see, for example, Freeman 1995). This paper demonstrates how the UCREL Semantic Annotation Scheme, a software program for automatic, dictionary-based content analysis, may be used to add a further dimension to both approaches, by systematically taking account of the semantic relationships between keywords via an investigation of key domains, and providing empirical support for some of the love-related conceptual metaphors put forward by cognitive metaphor theorists.

Specifically, we use the UCREL Semantic Annotation Scheme to explore the concept of *love* in three Shakespearean *love-tragedies* (*Othello*, *Anthony and Cleopatra* and *Romeo and Juliet*) and three Shakespearean *love-comedies* (*A Midsummer Night’s Dream*, *The Two Gentlemen of Verona* and *As You Like It*). Our approach is to initially determine how *love* is presented in the two datasets and then highlight any resemblances between our findings and the conceptual metaphors identified by cognitive metaphor theorists. We also discuss how the semantic field of *love* co-occurs with different domains in the two datasets, and assess the implications this has on our understanding of *love* as a concept. This research builds on Jonathan Culpeper’s work on keywords in Shakespeare, using Wordsmith Tools (Culpeper 2002); Paul Rayson’s comparisons of keyword and key domain analysis (Rayson 2003); and Dawn Archer and Paul Rayson’s work on the identification of key domains in refugee literature, using USAS (Archer and Rayson forthcoming).

Introduction

Keyword analysis has proven to be a very useful means of determining the aboutness of a text (or texts) and/or the *style* of a text, and for focussing researchers’ attention on aspects of a text (or texts) that deserve further enquiry. Importantly, a number of researchers who engage in keyword analysis group their keywords semantically, i.e. according to related or shared semantic space(s). For example, Baker (this volume) has grouped some of the keywords relating to a fox-hunting debate that took place in the British Parliament (e.g. *criminal*, *moral*, *barbaric*, *offence*, *absurd*, *illogical*, *rational*, *strong*, *tough*, *weakness*, etc.) in terms of their reaction to issues surrounding and/or their level of *cruelty*, *sensibleness*, *ethics* and *toughness*. Culpeper’s (2002) grouping of keywords relating to the main characters in *Romeo and Juliet* was determined by a different motivation, i.e. what they might tell us about characterization within the play. Romeo’s top three keywords – *beauty*, *blessed* and *love* – identify him as the lover of the play, for example, whilst other keywords relating to Romeo – *eyes*, *lips* and *hand* - highlight a related concern with the physical. Juliet, Romeo’s love interest, has very

different keywords, the most *key* being *if, yet, but* and *would*. On further investigation, Culpeper (2002: 20) has found that many of Juliet's usages of these lexical items 'reflect the fact that Juliet is in a state of anxiety for much of the play'. The keywords associated with Juliet's nurse differ from both Romeo and Juliet. Indeed, the majority – *god, warrant, faith, marry, ah* - can be categorised as surge features, that is to say, they reflect 'outbursts of emotion' (Taavitsainen 1999). Interestingly, when Culpeper (2002) explored these surge features further, he found that they marked occasions when the nurse was reacting to quite traumatic events (involving Juliet, in particular) and therefore should not be considered as a character trait, per se. This should alert us to the importance of contextualising keywords – a point often made but not always carried out convincingly.

In this paper, we take the grouping of keywords into related semantic spaces one step further by adopting a procedure that begins with the automatic identification of *key domains* in six Shakespearean plays – i.e. *Othello, Anthony and Cleopatra, Romeo and Juliet, A Midsummer Night's Dream, The Two Gentlemen of Verona* and *As You Like It* – using the UCREL Semantic Annotation System (USAS), and then goes on to identify *keywords* within these different *key domains*. The benefit of such an approach is that we are able to identify words that would not have been picked up by a keyword analysis (because they are not deemed to be *key* in and of themselves) but which nonetheless add to the aboutness of a text, because they share the same semantic space as the keywords. As will become clear, our approach also enables us to provide empirical support for the kinds of conceptual metaphor put forward by cognitive metaphor theorists when studying Shakespeare (see, for example, the work of Freeman (1995) and Barcelona Sánchez (1995)). By way of illustration, Barcelona Sánchez (1995) discusses the metaphorical basis of romantic love in *Romeo and Juliet*, in terms of the overarching concept metaphor, *love is the unity of its complimentary parts*.

As *love* is a common theme within Shakespeare and conceptual metaphors relating to *love* have been studied in some detail by cognitive metaphor theorists in a variety of literary and non-literary texts (including Shakespeare), we have chosen to explore the concept of *love* in our Shakespearean dataset.¹ However, rather than focussing on each text individually, *A Midsummer Night's Dream, The Two Gentlemen of Verona* and *As You Like It* are explored collectively as *love-comedies* and *Othello, Anthony and Cleopatra* and *Romeo and Juliet* are explored collectively as *love-tragedies*. The approach we adopt can be described as top-down, in that the categories are pre-defined and applied automatically by the USAS system. A breakdown of those categories and an explanation of the methodology we followed in this investigation are given in section 2. In section 3, we discuss the results of the automatic analysis, before moving on to an innovative study of key collocates at the domain level in section 4. Finally, section 5 concludes this paper by reflecting on the results and methodological implications of the research.

Methodology

In order to explore the key domains within our dataset, we initially annotated the *Nameless Shakespeare*, using USAS (Rayson et al, 2004), and made manual adjustments where necessary² before re-tagging the data in their collective groupings, i.e. *love-comedies* and *love-tragedies*. As the original USAS system was designed to undertake the automatic semantic analysis of present-day English language, we employed the historical version of the tagger in this investigation. The Historical Semantic Tagger, which has been developed by Archer and Rayson, includes supplementary historical dictionaries to reflect changes in meaning over time and a pre-processing step to detect variant (i.e. non-modern) spellings (Archer et al 2003). Indeed, the variant detector (VARD) tool currently searches for over 45,000 variant spellings and inserts the modern equivalent alongside the variant spelling in each case (as a 'reg tag').³ We have found that this feature, in particular, greatly facilitates the application of those standard corpus linguistic methods which are otherwise hindered by multiple variant spellings, i.e. frequency profiling, concordancing, key word analysis, etc. (see Rayson et al, 2005 for further details).

The taxonomy employed in the (modern and historical) USAS system presently uses a hierarchy of twenty one major domains, expanding into 232 semantic field tags. Table 1 shows the top level domains (see Appendix 1 for the full taxonomy):

Table 1. The Top Level of the USAS System

A. General and Abstract terms	B. The Body and the Individual	C. Arts and Crafts	E. Emotional Actions, States and Processes
F. Food and Farming	G. Government and the Public Domain	H. Architecture, Building, Houses and the Home	I. Money and Commerce
K. Entertainment, Sports and Games	L. Life and Living Things	M. Movement, Location, Travel and Transport	N. Numbers and Measurement
O. Substances, Materials, Objects and Equipment	P. Education	Q. Linguistic Actions, States and Processes	S. Social Actions, States and Processes
T. Time	W. The World and Our Environment	X. Psychological Actions, States and Processes	Y. Science and Technology
Z. Names and Grammatical Words			

The USAS system initially assigns part-of-speech tags to each word in a text prior to assigning one or more of the 232 semantic field labels. Portmanteau tags are used for those senses that straddle the borders of two or more semantic fields (such as *alehouse*, which borders the domains F and H). A key feature is the marking of multi-word expressions as single units for semantic analysis. The USAS taxonomy was derived from that of McArthur (1981), and has been considerably revised in the light of practical application. We are continuing to evaluate its suitability for the Early Modern English period through studies such as the one described in this paper.

The second stage of the analysis of *love* in the Shakespearean dataset was to compare semantic tag frequency profiles of the *love*-tragedies against the *love*-comedies. This was achieved using the log-likelihood statistic applied to the semantic tag frequencies. This step is analogous to the well-known key words procedure implemented in *WordSmith Tools* (Scott 2000). Here, we extended the technique to compare tag frequencies rather than word frequencies using the Wmatrix software (Rayson, 2003). By calculating the log-likelihood (LL) statistic for each tag and then sorting the profile by the result, we were able to see the most overused and underused semantic fields in the *love*-comedies relative to the *love*-tragedies. This technique has already been applied to a large *Forced Migration Online* corpus and has shown that improved efficiency over the standard key word technique can be achieved (Archer and Rayson, forthcoming).

The third stage of the analysis involved finding significant collocates for the key semantic fields. Our motivation was to discover which semantic tags collocate significantly with a small number of the key semantic tags selected from stage two. We were not aware of any off-the-shelf software tool which performs this task. We therefore used a Multilingual Corpus Toolkit (Piao et al 2002), which implements a number of well-known collocation statistics. The text was prepared from the tagged version by stripping out the words and leaving only the sequences of semantic tags and sentence breaks. The mutual information (MI) statistic with a window of +/-5 was applied to calculate tag collocates. One concern is that the relatively high frequencies of tags compared to words will result in negative MI values, but all the results we quote have positive MI values.

Results

In discussing our results below, we will make some reference to cognitive metaphor theory (as developed by the Lakoff, Johnson, Turner group). The application of cognitive metaphor theory to literary texts - Shakespeare in particular - has been established by people like Freeman (e.g. 1995). Given our interest in the concept of *love*, we will be making particular use of Barcelona Sánchez (1995), who analyses the love metaphors in *Romeo and Juliet* and Oncins-Martinez (forthcoming), who discusses metaphors relating to sexual activity in Early Modern English. However, the reader should note that our focus on conceptual metaphors in this paper is not meant to suggest that the USAS system can only be used for such analyses: indeed, Archer and McIntyre (2005) are presently using the USAS approach to investigate mind style in a number of literary texts (modern and

historical), and Archer and Rayson (forthcoming) have previously used the approach to investigate the different representations of refugees and their plight within refugee literature. Rather, our focus is merely meant to reflect the fact that some of the semantic fields we identify have metaphorical relationships with one another. That said, we do want to show how the USAS system may be used to provide empirical support for metaphorically-based research and, importantly, indicate previously-undiscovered conceptual metaphors. We will begin with a discussion of the most overused items in the comedies (relative to the tragedies).

The most overused items in the comedies relative to the tragedies

Nine semantic fields received a LL score above 15.13 in the comparison. This means that these semantic fields were significantly overused (at $p < 0.0001$ 1d.f.) in the *love*-comedies relative to the *love*-tragedies. As length constraints prevent a detailed discussion of all of the statistically significant fields, we limit our discussion in this section to the seven fields listed in Table 2 (following).

Table 2. The most overused items in the comedies relative to the tragedies

Semantic tag/field	Comedies		Tragedies		Log Likelihood
	Freq	%	Freq	%	
S3.2 = intimate/sexual relationship	379	.64	292	.36	55.50
L2 = living creatures	343	.58	279	.34	42.30
L3 = plants	149	.25	94	.12	35.99
S1.2.6- = (not) sensible	72	.12	32	.04	31.02
X3.1 = sensory: taste	120	.20	91	.11	18.41
E2+ = liking	325	.55	321	.39	17.36
T3- = old, new, young: age	153	.26	128	.16	17.12

It is noticeable that *S3.2 Intimate/sexual relationships* and *E2+ Liking* are amongst the most overused semantic fields in the *love*-comedies (relative to the *love*-tragedies). This means that *S3.2 Intimate/sexual relationships* and *E2+ Liking* represent two of the *most underused semantic fields* in the *love*-tragedies (when compared with the *love*-comedies). We will say more about this underuse of the *love*-related semantic fields in the *love*-tragedies in section 3.2.

Interestingly, the dominant lexical patterns within *S3.2 Intimate/sexual relationships* can be characterised in terms of Hallidayan-type (1994) participants (see Table 3) and processes (see Table 4):

Table 3. Participants in intimate/sexual relationships

<i>The two participants</i>	<i>Male participant</i>	<i>Male or female participant</i>	<i>Female participant</i>
Couples Lovers	Lover Suitor	Love	Virgin Wanton

Table 4. Processes in intimate/sexual relationships

<i>Transitive processes with male agents</i>	<i>Transitive processes with male or female agents</i>	<i>Intransitive processes with male or female agents</i>	<i>Transitive processes with female patients</i>
Kiss Kissing Kissed Kisses	Loves	Fall in love Falling in love Fallen in love Fell in love	Seduced deflowered

These results very clearly reflect an Early Modern patriarchal view of love, in which the male (in the role of *lover* or *suitor*) undertakes certain acts (e.g. *kissing*), which the female suffers (she is *seduced*, *deflowered*), with the result that she switches from *virgin* to *wanton*.

The semantic field *L2 Living creatures* appears as the second most overused field (see Table 2). Many of the lexical items within this field can be subsumed by the metaphor *love is a living being* and related metaphors, such as *the object of love is an animal*. Perhaps unexpectedly, and contrary to the items which Barcelona Sánchez (1995: 683) discusses for *Romeo and Juliet*, the bulk of these

items (e.g. *bears, serpent, snail, monster, adder, snake, claws, chameleon, worm, monkey, ape, weasel, toad, rat*) have strong negative associations, semantically speaking. Indeed, very few can be described as neutral (e.g. *cattle, horse, goats, creature* and *capon*) or positive (e.g. *deer, dove, nightingale*). Moreover, some of the items within the *positive* list are problematic; *deer*, for example, appears to be used positively in our texts. However, *deer* is linked to cuckoldry in many of Shakespeare's plays (e.g. *Love's Labours Lost, The Merry Wives of Windsor*) and, as such, may indicate that *deer* had a negative undertone for both Shakespeare and his audience¹. ⁴It is also worth noting that many of the negative lexical items in the *love*-comedies are personifications which relate to other metaphors suggested by Barcelona Sánchez (1995). They include *love is war*⁵ and *love is pain*, both of which are clearly relevant in the context of unrequited love, as the following example demonstrates:

Lysander Hang off , thou *cat* , thou burr ! vile thing, let loose ,
Or I will shake thee from me like a *serpent* !
Hermia Why are you grown so rude? what change is this?
Sweet love , —

(*A Midsummer Nights Dream*)

As will become clear, *love is war* and *love is pain* are also important conceptual metaphors in the *love*-tragedies (see section 3.2).

L3 Plants is the third most overused semantic field in the *love*-comedies (relative to the *love*-tragedies). This category is a useful one in that it highlights the importance of thoroughly checking the items captured by the different USAS tags. By way of illustration, the most frequent item in *L3 – mustardseed* – is a character's name, and the second most frequent item - *flower* – occurs as part of the multi-word unit *Cupid's flower*, i.e. the flower that Oberon used to send Titania to sleep in *A Midsummer Night's Dream*. More importantly, the bulk of the remaining items in the *Plants* semantic field can be explained by the fact that *As You Like It* and *A Midsummer Night's Dream* are set in woods (something which is not true of any of the tragedies) – and, if they were removed, the *keyness* of this category would probably decrease substantially. That said, a small number of the *L3* items have a strong metaphorical association with *love* or *sex*. For example, *Silvius* states that he is prepared to have *Phoebe* as a wife in spite of her less-than-virginal state in *As You Like It*.

Silvius So holy and so perfect is my love,
And I in such a poverty of grace,
That I shall think it a most plenteous *crop*
To glean the broken ears after the man
That the main *harvest reaps* : loose now and then
A scattered smile, and that I'll live upon.

According to Oncins-Martinez (forthcoming), this type of usage was common in the Early Modern English period. Indeed, he argues that the general conceptual metaphor *sex is agriculture* and its sub-mappings; *a woman's body is agricultural land, copulation is ploughing or sowing* and *gestation and birth is harvesting* underlie linguistic expressions which permeate many texts from this period.

The semantic field *S1.2.6 (Not) sensible* does not refer to the older meaning of 'not having the capacity to sense (feel)' but to being foolish, silly, stupid, and so on. Interestingly, the metaphorical associations are much stronger in this semantic field than in *L3 Plants*. Indeed, many of the items can be accounted for by Lakoff and Johnson's (1980: 49) *love is madness* metaphor (cf. Barcelona Sánchez's (1995: 679) *love is insanity*), as exemplified here:

Silvius O Corin, that thou knew how I do love her!
Corin I partly guess; for I have loved ere now.
Silvius No, Corin, being old, thou
Can not guess,

Though in thy youth thou was as true a lover
 As ever sighed upon a midnight pillow:
 But if thy love were ever like to mine—;
 As sure I think did never man love so—;
 How many actions most *ridiculous* have thou been drawn to by thy fantasy?
 Into a thousand that I have forgotten.
 O, thou did then ne'er love so heartily!
 If thou remember not the slightest *folly*
 That ever love did make thee run into,
 Thou have not loved [...]

(As You Like It)

T3 Old, new, young: age also appears to relate to the *love is madness* metaphor, for the reason that being young and being in love are assumed to be states that are accompanied by a lack of rational thought. In the following extract from *A Midsummer Night's Dream*, for example, Love is 'said to be a child' because of its capacity to *beguile*:

Nor has Love's mind of any judgement taste;
 Wings and no eyes figure unheedy haste:
 And therefore is Love said to be a *child*,
 Because in choice he is so oft beguiled.

Significantly, all the items within this semantic field relate to the early years of life, the most frequent items being *youth* and *young*. Some of these lexical items, in turn, modify *lover(s)*.

The lexical items that constitute X3.1 *Sensory: taste* fall into three groups:

sweet	Bitter	taste
sweetest	bitterness	tastes
sweeter	sourest	
	sour	

The first group is very much part of *sweet talk* used in courtship. The most frequent item by far is *sweet* with 94 instances⁶; the next most frequent item being *bitter* with 12 instances. The connection with love can be seen in the metaphor *love is food* (Barcelona Sánchez 1995: 672-3). In the example below, *loving words* become *sweet honey*:

Julia Nay, would I were so angered with the same!
 O hateful hands, to tear such loving words!
 Injurious wasps, to feed on such *sweet* honey
 And kill the bees that yield it with your stings!

(Two Gentlemen of Verona)

Sweet often appears as part of vocative expressions, as in *sweet lady*. Although it is used for men and women together (e.g. *sweet lovers*) and for men (e.g. *sweet Proteus*), the vast majority of the instances refer to women. A possible explanation for this lies in the metaphor *a woman's body is agricultural land*, which we have discussed briefly in relation to *L3 Plants* (see above). As Oncins-Martinez (forthcoming) explains, the land gives rise to trees, metaphorically representing people, and the trees give rise to fruits, metonymically associated with a woman's sexual attributes. Then a link can be established with *woman is an edible substance* and *sex is eating*. Hence, referring to a woman as *sweet* can be seen as instantiating metaphors relating to women as edible sexual objects. The second group of items in this semantic field, relating to bitter/sour, often relate to the troubles of love (e.g. unrequited love).

Rosalind He's fallen in love with your foulness and she'll fall in love with my anger. If it be so, as fast as she answers thee with frowning looks. I'll sauce her with *bitter* words.
(As You Like It)

Note how, in this particular example, the word *sauce* relates to the *love is food* metaphor: anger leads to love which is food; thus anger will be a bitter sauce. The OED lists this particular example as an illustration of the sense 'to rebuke smartly', but the earlier sense 'to season, dress, or prepare (food) with sauces or condiments' was still current.

We have left our discussion of *E2+ Liking* until now as our investigations have shown that it is not currently a secure category. For example, the most frequent item *like* is almost always a preposition, and we are interested in *like* as a verb. Moreover, cases like *loved*, *loving*, *beloved*, *dotes*, *enamoured*, *adores*, *adored*, *adoration*, *amorous*, and *doting* are not always dealt with on a sufficiently principled basis (that is to say, different variants of the base form of these particular items occur in different positions in the categorization of both *E2+ Liking* and *S.3.2 Intimate/sexual relationship*). It is perhaps not surprising that these categories closely inter-mesh, since:

- *liking* stands in a very close relationship with *intimate/sexual relationship*;
- intimate or sexual relationships presuppose physical closeness;
- physical closeness can be caused by love (a metonymic *effect for cause* relationship, cf. Barcelona Sánchez 1995: 671);
- and *liking* stands in a metonymic relationship with *love* ('a part of love stand[s] for the whole concept of love': Barcelona Sánchez 1995: 675).

However, currently English Modern words may not accurately be mapped onto such a semantic network of relations. For example, *lover* did not necessarily indicate a physically intimate relationship in the Early Modern English period, but, rather, could also mean *friend* (Crystal and Crystal 2005).

In Table 5, we have combined the *E2+ Liking* and *S.3.2 Intimate/sexual relationship* lexical items for the *love-comedies* and the *love-tragedies* together in a *love-related* macro category. We have also italicized the lexical items that only occur in the comedies or only occur in the tragedies, so that we can more readily highlight the stronger negative associations inherent in some of the lexical items relating to the latter (e.g. *bewhored*, *carnal*, *cuckold* and *sluttish*). Notice that these particular items appear to provide empirical support for Barcelona Sánchez's (1995: 684) claim that the concept of tragical love is 'characterized by being *adulterous* [love] inevitably ending in death':

Table 5. Love-related lexical items which occur in the comedies and tragedies

Comedies	Tragedies
Adoration (1), adore (2), adored (1), affection (7), affections (1), <i>after-love</i> (1), amorous (1), applaud (2), applause (1), <i>apple of his eye</i> (1), beloved (8), chastity (3), cherish (1), cherished (2), <i>copulation</i> (1), <i>couples</i> (4), dear (32), deflowered (1), <i>dote</i> (8), dotes (4), doting (1), enamoured (2), enjoy (2), fall in love (4), fallen in love (1), falling in love (1), fancies (1), fancy (7), fell in love (1), fond (7), <i>gone for</i> (1), impress (1), <i>in love</i> (34), kiss (20), kissed (3), kisses (2), kissing (3), like (117), liked (2), liking (2), <i>love</i> (354), loved (28), <i>lover</i> (32), <i>lovers</i> (26), loves (26), loving (10), paramour (2), precious (5), prized (1), relish (2), revelling (1), revels (5), <i>savours</i> (3), <i>seduced</i> (1), <i>sensual</i> (1), suitor (2), take to (1), that way (1), virgin (4), wanton (5)	adore (1), adores (1), affection (8), affections (7), <i>affinity</i> (1), amorous (6), applauding (1), applause (1), beloved (5), believing (1), <i>bewhored</i> (1), <i>carnal</i> (1), <i>chamberers</i> (1), chastity (3), cherish (1), cherished (1), cherishing (1), courts (1), <i>cuckold</i> (6), <i>darling</i> (1), <i>darlings</i> (1), dear (59), deflowered (1), <i>devotion</i> (4), dote (1), dotes (2), doting (5), enamoured (1), enjoy (3), enjoyed (1), fall in love (1), fancies (2), fancy (4), fond (8), impress (1), in love (8), kiss (25), kissed (6), kisses (13), kissing (6), liked (1), likes (1), liking (1), likings (1), love (259), loved (21), lover (6), lovers (9), loves (29), loving (12), <i>lust</i> (10), <i>lusts</i> (1), paramour (1), precious (6), prized (1), rate (3), rated (1), relish (1), revel (3), revels (4), <i>sluttish</i> (1), suitor (2), suitors (2), take to (1), that way (1), <i>the other way</i> (2), wantons (1), <i>wooer</i> (1)

Significantly, several of these love-related lexical items are *keywords* in and of themselves in the *love-comedies* relative to the *love-tragedies*. Indeed, *love*, *in love* and *lover* all have LL scores above our cut-off point of 15.13, whilst *lovers*, *dote*, *betrothed*, *couples* and *virgin* have LL scores between 6.91 and 15.13 (see italicised items). Interestingly, the USAS system chose to assign three key words with a potential link to love - *betrothed*, *woo* and *desire* – to other semantic fields when assigning LL

scores. The reader should be aware, then, that a key word and key domain analysis of the same data will reveal both overlap and difference (see Hoover et al (forthcoming) for a more detailed discussion of this potential overlap/difference). As Table 5 reveals, the strength of the USAS system is that it can identify words that would not have been picked up by a keyword analysis (because they are not deemed to be *key* in and of themselves) but which nonetheless add to the aboutness of a text, as they share the same semantic space as the keywords. However, as the USAS process is an automatic one, it is important that any results are checked thoroughly to determine their contextual relevance. That said, we would point out that the keyword list showing overused items in the *love-comedies* (relative to the *love-tragedies*) stretches to some 275 items in comparison to 9 key domains, and that the entries in the keyword list are ambiguous for part-of-speech and sense. Consequently, a manual examination of concordance lines is required in addition to manual grouping into semantic patterns (see Rayson 2003: 100-113 for a more detailed exploration of the advantages of the key domains approach).

The most overused items in the tragedies (relative to the comedies)

Twelve semantic fields from the tragedies achieved an LL score of over 15.13 indicating significance at the 99.99% level ($p < 0.0001$ 1 d.f.). Due to length constraints, we will be concentrating on seven specific semantic fields in the *love-tragedies* relative to the *love-comedies* (see Table 6):

Table 6. Most overused items in the tragedies relative to the comedies

Semantic tag (field)	Comedies		Tragedies		Log Likelihood
	Freq	%	Freq	%	
G3 = warfare, defence, and the army	425	.52	57	.10	213.51
L1- = (lack of) life/living things	490	.60	170	.29	77.16
Z2 = geographical names	399	.49	153	.26	49.56
E3- = (not) calm/violent/angry	343	.42	143	.24	33.67
M4 = movement (by sea/through water)	92	.11	21	.04	28.51
S9 = religion and the supernatural	644	.79	345	.58	21.92
S7.1- = (lack of) power/organising	193	.24	77	.13	21.55

Notice the lack of semantic fields directly to do with love (e.g. *Intimacy/sexual relationships, Liking*), and the appearance of fields, such as *warfare, lack of life or living things* and *geographical names* that seem to have nothing to do with love. As previously explained, this is because the love-related semantic fields (*E2+ Liking* and *S3.2 Intimate/sexual relationships*) are actually amongst the most underused categories in the *love-tragedies* relative to the *love-comedies*. However, it is worth noting that many of the categories that are overused do have metaphorical links with love, albeit to differing degrees. By way of illustration, some of the items within *G3 Warfare, defence, and the army* reflect the *love is war* metaphor (cf. Barcelona Sánchez 1995: 678-9), briefly mentioned above. Romeo, for example, comments that 'She will not stay the *siege* of loving terms, Nor bide th' encounter of assailing eyes'. Some items in the *G3* category have nothing to do with love, of course, capturing, instead, other aspects of the tragedy: The most frequent of these items – *Soldier* – occurs most frequently in *Anthony and Cleopatra*, as do most cases of *sword, war, wars, army, battle, armour* and *navy*. This is not surprising: *Anthony and Cleopatra* involves military power struggles. There are a few cases from *Othello*: Othello, and the other men, are military folk, with a military history. *General* and *lieutenant* are vocative forms which nearly always refer to Othello and Cassio respectively. There are very few items from *Romeo and Juliet*: indeed, 50% of the occurrences of *swords* occur in *Romeo and Juliet* (in the fight between Benvolio and Tybalt), as do most of the occurrences of *dagger* (in the fight between Mercutio and Tybalt, and Juliet's suicide with Romeo's dagger).

This most key semantic field in the *love-tragedies* parallels the most key field in the *love-comedies* - *S3.2 Intimate/sexual relationship* – to some extent, in that both capture the distinctive participants and processes that characterise the plots of the two genres. However, the three *love-tragedies* are not equally characterised by literal warfare. Indeed, the warfare in *Romeo and Juliet* tends to be metaphorical. Significantly, where there is literal warfare (or military activity), there is also

a strong link with the sea (which helps to explain the keyness of the *M4 Movement by sea through water* domain in the *love-tragedies*).

The 15 most frequent items in the semantic field *L1 (Lack of) life/living things* are: *death, dead, die, kill, slain, murder, dies, killed, mortal, tomb, dying, murdered, corpse, fatal, drowned*. Not surprisingly, the bulk of these appear at the ends of the plays; in the death scenes. The field contains a mixture of literal and metaphorical usages. In the following example, *death* is personified as Juliet's lover:

Ah, dear Juliet, Why art thou yet so fair? shall I believe
That unsubstantial *death* is amorous,
And that the lean abhorred monster keeps
Thee here in dark to be his paramour?

(*Romeo and Juliet*)

As Barcelona Sánchez (1995: 684) points out, the concept of tragical love is 'characterized by being *adulterous* inevitably ending in death'.

The *Z2 Geographical names* field contains a number of miscategorizations (e.g. *Moor*), and so its appearance as a highly key semantic field must be treated with caution. Nevertheless, this field does reflect the fact that some of the plots of the *love-tragedies* involve a number of different geographical locations. This is particularly true of *Anthony and Cleopatra*, in which the action flips back and forth between Rome and Egypt. This field seems to have no obvious link with love, metaphorical or otherwise, however.

Generally, the category *E3- (Not) calm/violent/angry* captures the violent conflicts that characterise the tragedies. *Poison* is the most frequent item. It may appear rather odd in this category, but it is presumably designed to capture the sense of someone who is strongly hated (cf. OED 3.b). Of course, this is a modern sense that does not apply to our data. Indeed, most of our instances relate to the literal poison in *Romeo and Juliet*. However, the idea that *love is poison* is also articulated in the plays, as the following example, where Cleopatra talks about Anthony, makes clear:

He 's speaking now ,
Or murmuring "Where's my serpent of old Nile?";
For so he calls me: now I feed myself
With most delicious *poison*. Think on me,
That am with Phoebus amorous pinches black,
And wrinkled deep in time?

(*Anthony & Cleopatra*)

Some *E3- (Not) calm/violent/angry* items, i.e. *angry, rage* and *fury*, occur in all the plays. Interestingly, although there is no obvious *metaphorical* link with love, there is a link with the more negative aspects of love, such as adultery, jealousy and revenge. These *emotional states* arise mainly as a consequence of the tragic plot. By way of illustration, as Othello's jealousies spiral at the presumed adultery of his wife Desdemona, Iago innocently asks 'Is my lord *angry*?', whilst Desdemona, perplexed, asks 'What, is he *angry*?'. All but one of the nine instances of *revenge* come from *Othello*, the only play in our data that has the plot of a Revenge Tragedy'.

The results of the semantic field *S9 Religion and the supernatural* are slightly skewed by 85 instances of *Friar* as a term of address (and also by *holy* in the vocative *Holy Friar*). Similarly, *pray* is usually part of the politeness formula 'I pray you'. The most frequent item - *heaven* - is often used as part of an appeal (e.g. '*heaven* defend your good souls', *Othello*). Almost all of the uses of *heaven* are from *Othello* and, to a lesser extent, *Romeo and Juliet*. The same is also true of the distribution of *soul* and *devil*. This result may be indicative of the characters' experiences in these plays, in the sense that some of the characters endure more protracted bouts of torment as a consequence of the tragic plot. Alternatively, some of the usages of *heaven, soul* and *devil* within *Othello* may reflect the fact that Othello is a deeply religious character.

A small number of the *S9 Religion and the supernatural* items can be accounted for by the metaphor *object of love is a deity* (Barcelona Sánchez 1995: 674), as illustrated by the following examples:

Roderigo I can not believe that in her; she's full of most *blessed* condition. (*Othello*)

Romeo If I profane with my unworhiest hand This *holy shrine*, the gentle sin is this:
My lips, two blushing *pilgrims*, ready stand
To smooth that rough touch with a tender kiss.
(*Romeo and Juliet*)

Rather fewer of the items in this category appear in *Anthony and Cleopatra*, but, when they do, they tend to reflect the non-Christian setting of the play (i.e. *gods* and *soothsayer*).

Our final semantic field, *S7.1- (Lack of) power/organising*, is somewhat skewed by the most frequent item - *servant* - which is a character name (e.g. 'First Servant'). The item *wench* is a another possible miscategorization, as the original meaning of 'young woman' was still current in this period, alongside newer senses indicating a 'girl of the rustic or working class', 'a wanton' or 'a female servant'(OED) – senses which indicate someone of low power. But there are many items – *knave*, *sirrah*, *minion*, *churl*, *slave* and so on – that clearly do reflect lack of power, and are often used to hurl abuse at characters. The bulk of these do not connect with love, or even the tragic conception of it. Instead, they reflect the fact that the *love*-tragedies revolve around hierarchical power structures rather more than the *love*-comedies.

Moving towards an analysis of collocations at the domain level

As well as being interested in key domains, we are interested in the extent to which important collocate information can be discovered at the domain level (rather than the word level). Table 7, then, captures the domains that the category *S3.2 Intimate relationship* collocated most strongly with in the *love*-comedies. Due to length constraints, our discussion here will concentrate on two of the four: *B1 Anatomy & physiology* and *Z8m Pronouns*.

Table 7. Domain collocates of *S3.2 Intimate relationship* in the comedies

O2= objects	MI =2.392	'I kiss the <i>instrument</i> of their pleasures'
A1.1.1 = general actions	MI =1.412	'Think true love <i>acted</i> simple modesty'
B1 = anatomy & physiology	MI =1.317	'... a fire sparkling in lovers <i>eyes</i> '
Z8m = pronouns (male)	MI =1.298	'if thou Can cuckold <i>him</i> , thou do thyself a pleasure, me a sport'

The presence of the semantic field *B1 Anatomy & physiology* is not surprising, given that the 'embodiment' of meaning is perhaps the central idea of the cognitive view of meaning (Kövecses 2002: 16). Moreover, the human body, so close to us and tangible, is an obvious source domain for metaphorically understanding abstract targets such as love. *Eyes* (or *eye*) and *heart* are the most frequent items within the *B1* semantic field, with 175 occurrences. The bulk of the instances of *eyes* occur in *A Midsummer Night's Dream* – remember that Puck puts the love potion in Titania's eyes.

Elsewhere, there is a strong notion that a woman's eyes were an aspect of her beauty that could capture men. Barcelona Sánchez (1995: 679) suggests that the underlying metaphor here is *eyes are containers for superficial love*, which seems to be a development of Lakoff and Johnson's *eyes are containers for the emotions*. In fact, the idea of a container is not clearly articulated in the comedy data. Indeed, we would suggest that *eyes are weapons of entrapment* is a more appropriate conceptual metaphor for our data, as here (cf. the related metaphor *love is war*):

Orlando Wounded it is, but with the *eyes* of a lady.
(*As You Like It*)

Valentine This is the gentleman I told your ladyship
Had come along with me, but that his mistress
Did hold his *eyes* locked in her crystal looks.

Silvia Belike that now she has enfranchised them
 Upon some other pawn for fealty.
Valentine Nay, sure, I think she holds them prisoners still.
Silvia Nay, then he should be blind; and, being blind
 How could he see his way to seek out you?
Valentine Why, lady, Love has twenty pair of eyes.
Thurio They say that Love has not an eye at all.
Valentine To see such lovers, Thurio, as yourself:
 Upon a homely object Love can wink.

(*Two Gentlemen of Verona*)

Helena How happy some o'er other some can be!
 Through Athens I am thought as fair as she.
 But what of that? Demetrius thinks not so;
 He will not know what all but he do know:
 And as he errs, doting on Hermia's eyes,
 So I, admiring of his qualities:
 Things base and vile, holding no quantity,
 Love can transpose to form and dignity:
 Love looks not with the eyes, but with the mind;
 And therefore is winged Cupid painted blind:
 Nor has Love's mind of any judgement taste;
 Wings and no eyes figure unheedy haste:
 And therefore is Love said to be a child,
 Because in choice he is so oft beguiled.
 As waggish boys in game themselves forswear,
 So the boy Love is perjured every where:
 For ere Demetrius looked on Hermia's eyne,
 He hailed down oaths that he was only mine [...]

(*A Midsummer Night's Dream*)

Apart from the literal sense, a *heart* can stand for a person, via a part-whole metonymic relationship. They can be further understood as containers for love, the metaphor being *the heart is a container of emotions* (Barcelona Sánchez 1995: 670). Thus, hearts can harden (e.g. 'if your *heart* be so obdurate, Vouchsafe me yet your picture for my love', *TGV*), not allowing access to the emotional reservoir inside, or the protective container can be pierced, damaging the emotional contents (e.g. 'Pierced through the *heart* with your stern cruelty', *A Midsummer Night's Dream*). Expressions such as 'with all my heart' seem to have the sense: 'with all the emotions within my heart'. Hearts, or more accurately the emotions within, can also be attributed agency (e.g. 'Here is her hand, the agent of her *heart*', *TGV*), and can be personified (e.g. 'My *heart* to her but as guest-wise sojourned, And now to Helen is it home returned, There to remain', *A Midsummer Night's Dream*).

Tears are closely tied to unrequited love, as the following example makes clear:

Phoebe Good shepherd, tell this youth what it is to love .
Silvius It is to be all made of sighs and tears;
 And so am I for Phoebe .

(*As You Like It*)

Tears have a cause-effect metonymic relationship with pain or emotional distress: they are the effect, and pain is metaphorically related to love (*love is pain*). Consider the following in which love personified inflicts pain with resultant tears and sighs:

I have done penance for contemning Love,
 Whose high imperious thoughts have punished me

With bitter fasts, with penitential groans,
 With nightly *tears* and daily heart-sore sighs.

(*Two Gentlemen of Verona*)

The collocation between *Z8m pronouns (male)* and *S3.2 Intimate relationship* in the *love-comedies* is not surprising when one considers the extent to which female characters - in particular, Julia and Rosalind - talk about the men they love (e.g. Proteus and Orlando respectively). Significantly, other female characters within the *love-comedies* appear to share their affinity for male pronouns, as the following extract taken from *As You Like It* demonstrates:

Phoebe Think not I love *him*, though I ask for *him*; it is but a peevish boy; yet *he* talks well;
 But what care I for words? yet words do well
 When *he* that speaks them pleases those that hear.
 It is a pretty youth: not very pretty:
 But, sure, *he's* proud, and yet *his* pride becomes *him*:
He'll make a proper man: the best thing in *him*
 Is *his* complexion; and faster than *his* tongue
 Did make offence *his* eye did heal it up.
He is not very tall; yet for *his* years *he* 's tall:
His leg is but so so; and yet it is well:
 There was a pretty redness in his lip,
 A little riper and more lusty red
 Than that mixed in *his* cheek; it was just the difference
 Betwixt the constant red and mingled damask.
 There be some women, Silvius, had they marked *him*
 In parcels as I did, would have gone near
 To fall in love with *him*; but, for my part,
 I love *him* not nor hate *him* not; and yet have more cause to hate *him* than to love *him*:
 For what had *he* to do to chide at me?
He said mine eyes were black and my hair [...]

As Table 6 highlights, gender pronouns are also a key collocates of the *love-tragedies*. However, it is the female pronoun that is key here, which suggests that, rather than women characters talking about their love for a man *directly* (as occurs in the *love-comedies*), male and female characters in the tragedies are *reporting* a female's love for a man (cf. the *discussions* respecting Desdemona's love for Othello, Cleopatra's love for Anthony and, to a lesser extent, Juliet's love for Romeo).

Table 8. Domain collocates of S3.2 Intimate relationship in the tragedies

N3.2+/A2.1 = change in size	MI =5.014	'But my true love is <i>grown</i> to such excess'
A5.2+ = evaluation ('true')	MI =2.774	'For if he be not one that <i>truly</i> loves you'
M2 = movement/transporting	MI =2.099	'Look, if my gentle love be not <i>raised up!</i> '
S6+ = obligation & necessity	MI =1.944	'I must show out a flag and sign of love'
Z8f = pronouns (female)	MI =1.844	'It can not be long that Desdemona should continue her love to the Moor'
A7+ = certainty	MI =1.604	'... if thou Can cuckold him, thou do thyself a pleasure, me a sport'
A1.1.1 = general actions	MI =1.354	'O mistress, villainy has made mocks with love!'

Summary of main findings

In this paper, we have reported on an exploration of key domains within three Shakespearean *love*-comedies and three Shakespearean *love*-tragedies. We have observed marked differences in the occurrence of *love* in our two datasets. This is clearly represented by the semantic fields of *intimate/sexual relationships* and *liking* appearing as the most underused concepts in the *love*-tragedies when compared to the *love*-comedies. The *love*-tragedies focus, instead, on *war, lack of life/living things, religion and the supernatural, lack of power, movement*, etc., some of which highlight interesting metaphorical patterns. We have also observed that, when love is represented in the *love*-tragedies, it is much 'darker', and, as such, may typify the 'tragical' love (as opposed to 'ideal' or 'romantic' love) identified by Kövecses (1986) (see also Barcelona Sánchez 1995). Many of our results have been explained in terms of cognitive metaphor theory. This is not surprising, as abstract concepts such as love are difficult to express, and so metaphor is used. However, it should be noted that key domain analysis is not only concerned with the identification of metaphorical patterns. Indeed, in *Anthony and Cleopatra* in particular, the key semantic fields tended to identify (or relate to) the tragic plot.

As a result of this study, we plan to refine several USAS categories, so that we can capture differences in pronoun usage more readily, e.g. in respect to gender and subject/object positioning. We are also actively investigating the inclusion of further components within the Wmatrix structure, which will allow us to distinguish metaphorical usage and to calculate domain collocation statistics.

Concluding comments

We have shown that the analysis of key domains is a useful methodology in that it enables us to discover links across different semantic fields that may not become readily apparent when using a key words analysis or analysing texts manually. Key domains also provide a way in to cognitive metaphor-type analysis in that we can identify lexical semantic patterns using the USAS system, which, upon closer manual inspection, may be found to be metaphorically linked to particular domains. We believe that one of the greatest strengths of the USAS system is that we are able to compare huge amounts of data in a relatively short amount of time; however, the analyst must always keep in mind the limits of automatic annotation tools. As such, we advocate that quantitative analysis is always combined with qualitative analysis.

Acknowledgements

The work presented in this paper was carried out within two projects: *Unlocking the Word Hoard* funded by the Andrew W. Mellon Foundation with Martin Mueller of Northwestern University and *Scragg Revisited* funded by the British Academy (small research grant scheme SG-40246).

References

- Archer, D. and Rayson, P., 'Using the UCREL automated semantic analysis system to investigate differing concerns in refugee literature', In M. Deegan, L. Hunyadi, and H. Short (eds.) *The Keyword Project: Unlocking Content Through Computational Linguistics*. (London: Office for Humanities Communication Publications, forthcoming), 18.
- Archer, D. and McIntyre, D., 'A computational approach to mind style', Paper given at the 25th Conference of the Poetics and Linguistics Association, University of Huddersfield, UK, July 2005.
- Archer, D., McEnery, T., Rayson, P., Hardie, A., 'Developing an automated semantic analysis system for Early Modern English', In D. Archer, P. Rayson, A. Wilson and T. McEnery (eds.) *Proceedings of the Corpus Linguistics 2003 Conference. UCREL technical paper number 16* (Lancaster: UCREL, 2003), 22–31.
- Barcelona Sánchez, A., 'Metaphorical models of romantic love in *Romeo and Juliet*', *Journal of Pragmatics*, 24 (1995), 667–688.
- Crystal, D. and Crystal, B. *The Shakespeare Miscellany* (London: Penguin, 2005).

Culpeper, J., 'Computers, language and characterisation: An analysis of six characters in *Romeo and Juliet*', in Ulla Melander Marttala, Carin Ostman and Merja Kyto (eds), *Conversation in Life and in Literature: Papers from the ASLA Symposium* (Uppsala: Association Suedoise de Linguistique Appliquee, 2002), 11–30.

Freeman, D.C., "'Catch[ing] the nearest way": *Macbeth* and cognitive metaphor', *Journal of Pragmatics*, 24 (1995), 689–708.

Halliday, M.A.K., *An Introduction to Functional Grammar*. (2nd edition, London: Edward Arnold, 1994).

Hoover, D., Culpeper, J. and Louw, B., *Approaches to Corpus Stylistics* (London: Routledge, forthcoming)

Kövecses, Z., *Metaphor: A Practical Introduction* (New York: OUP, 2002)

Lakoff, G. and Johnson, M., *Metaphors we Live by* (Chicago and New York: The University of Chicago Press, 1980)

McArthur, T., *Longman Lexicon of Contemporary English* (Longman: London, 1981)

Oncins-Martinez, J.L., *Notes on the Metaphorical Basis of Sexual Language in Early Modern English* (forthcoming).

Piao, S.L., Wilson, A. and McEnery, T., 'A Multilingual Corpus Toolkit', paper given at AAACL-2002, Indianapolis, Indiana, USA, 2002.

Rayson, P., 'Matrix: A Statistical Method and Software Tool for Linguistic Analysis Through Corpus Comparison'. Ph.D. thesis (Lancaster University, 2003).

Rayson, P., Archer, D. and Smith, N., 'VARD versus Word: A comparison of the UCREL variant detector and modern spell checkers on English historical corpora', *Proceedings of the Corpus Linguistics Conference Series On-Line E-Journal*, 1:1, 2005.

Rayson, P., Archer, D., Piao, S. L., McEnery, T., 'The UCREL semantic analysis system', *Proceedings of the Workshop on Beyond Named Entity Recognition Semantic labelling for NLP Tasks in Association with the Fourth International Conference on Language Resources and Evaluation (LREC, 2004)*, 7–12.

Scott, M. 'Focusing on the Text and Its Key Words', in L. Burnard & T. McEnery (eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (Frankfurt: Peter Lang, 2000), 2, 103–122.

Taavitsainen, I., 'Personality and Styles of Affect in *The Canterbury Tales*', in G. Lester (ed.), *Chaucer in perspective: Middle English Essays in Honour of Norman Blake*, (Sheffield: Sheffield Academic Press, 1999), 218–234.

Scott, M. 'Focusing on the Text and Its Key Words', in L. Burnard & T. McEnery (eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (Frankfurt: Peter Lang, 2000), 2, 103–122.

Taavitsainen, I., 'Personality and Styles of Affect in *The Canterbury Tales*', in G. Lester (ed.), *Chaucer in perspective: Middle English Essays in Honour of Norman Blake*, (Sheffield: Sheffield Academic Press, 1999), 218–234.

Notes

¹ The source data for our research is taken from the *Nameless Shakespeare Corpus* hosted by Northwestern University. For more details, see <<http://www.library.northwestern.edu/shakespeare>>

² We were aided in the initial tagging and checking process by students from Northwestern University.

³ Dawn Archer, Paul Rayson and Nick Smith are presently developing VARD2, which will make use of context rules and fuzzy matching algorithms as well as the simple search and replace script. The context rules will allow the detection of real-word spelling variants (for example *bee* instead of *be*) and the detection (and 'correction') of morphological inconsistencies (for example, the 'correction' of (e)s to 's' where we would expect the genitive today). The fuzzy matching algorithms will allow matching of previously unseen variants to their 'correct' modern equivalents.

⁴ Prof. John Joughin, personal comment.

⁵ *love is war* is a long-established metaphor for love (see, for example, Lakoff and Johnson 1980: 49).

⁶ *Sweet* also occurs as a key word, with a LL score of 19.14.