

Identifying Multi-Word Units in Context

Gill Philip

Università degli Studi di Bologna, Italy

Università degli Studi di Bologna

CILTA

Piazza S. Giovanni in Monte, 4

40124 Bologna

Italy

tel. +39 051 209 7763

fax. +39 051 209 7751

e-mail g.philip@cilta.unibo.it

Abstract

Far from being linguistic anomalies, multi-word expressions abound in natural language, yet their identification is surprisingly problematic. The same combination of words can occur as a compositional, fully lexical string or as a delexicalised multi-word unit (MWU). How can these different manifestations of a series of words be distinguished one from the other? To exacerbate the problem, the creativity of language users results in the appearance of non-canonical forms of MWUs. How can these innovative uses be retrieved so that they can be incorporated into a comprehensive analysis of the MWU under study? This paper sets forth procedures for retrieving non-canonical variants from large general reference corpora, and addresses the disambiguation of compositional and non-compositional multi-word strings from a collocational standpoint.

Keywords: canonical, variation; non-compositional; salience

Introduction

When does a string of words become identifiable as a multi-word unit? In theoretical terms, the distinction is quite straightforward: normal word combinations can be decomposed into their component parts, whereas multi-word units cannot. So, for example, *black sheep* (a sheep with a black fleece) is compositional, but *black sheep [of the family]* (a person whose behaviour deviates from the accepted norm) is not used to talk about sheep of any colour; it is non-compositional and idiomatic. The simplicity of such ‘textbook’ examples, and the clear-cut distinctions that they present, belie the complexities of multi-word expressions as they are manifested in language use. The traditionally-favoured canonical forms constitute neat citational examples which provide clear illustrations of the theoretical distinctions being made; but by illustrating only the extremes of the compositional – non-compositional continuum, they oversimplify a very complex area of language and are unable to address the swathe of language that occupies the grey area in between.

Corpora not only demonstrate that non-canonical forms abound in language; they also allow these forms to be analysed and classified (see Moon, 1998; Philip 2003). But there is a significant hurdle to be overcome in that the corpus can only output concordance lines which match the search query entered by the linguist. Typically, a search query consists of the word or collocation under study, and the data is progressively refined and restricted as the analysis proceeds. In the case of variant forms of multi-word units, however, the very key word or collocation under study may themselves be altered, resulting in a much more vague and ill-formed definition of the object of study. Should the linguist wish to find non-canonical forms of a multi-word unit, he or she must therefore formulate a query, or series of queries, which allows as many of them to be retrieved as possible, without necessarily being in possession of any *a priori* knowledge or assumptions regarding their precise lexical realisations.

The inclusion of variant forms in the analysis of multi-word expressions such as idioms and metaphorical phrases is of substantial importance, as the identification of a language chunk as a non-compositional whole stands in opposition to the identification of the same lexical string in a compositional guise, for example in the expression *black sheep*. In order to define what a multi-word unit is, it is necessary to define what it is not. Yet this area of study tends to be overlooked; it is generally taken for granted that compositional language is the norm, whereas non-compositional chunks are somehow anomalous, in much the same way as ‘literal’ language is contrasted with the myriad definitions and classifications of ‘figurative’ language (Gibbs 2002:474). The question that will be addressed here is therefore: in what way(s) are compositional collocations significantly different to non-compositional phrases, and how can these differences be quantified?

The data used in this paper is taken from the Bank of English (University of Birmingham/HarperCollins publishers).

1. Compositional and non-compositional collocations

The recognition of a word string as non-compositional is core to its identification as a multi-word unit rather than simply a fortunate co-occurrence of words. A collocation

may well be fixed, such as *black and white* (as opposed to **white and black*), but this in itself does not make the unit non-compositional. Of central importance to the recognition of the string as a holistic unit is not the words themselves, but what those words mean when they appear in combination. *Black and white* is simply the accepted order for these colours to be listed, irrespective of whether or not the colours are intended in a literal or a metaphorical sense, and the collocation can be decomposed into its component parts. On the other hand, the same collocation becomes holistic when its combined meaning takes on a distinct sense which is no longer directly related to black things and white things, as is the case with the idiomatic sense of *black and white*, which can be glossed as *clear-cut*. As decontextualised citational forms, these two senses of the same lexical string cannot be disambiguated, but in context, their intended meaning becomes apparent. It must therefore be concluded that the single most important disambiguating factor is the context in which the string occurs, rather than its internal make-up.

Linguists in the neo-Firthian tradition break context down into two complementary aspects which enable the analyst to separate out the social and pragmatic reality in which the language is uttered – the *context of situation* – and the words which occur together with the item under study – the *cotext*. Although extra-linguistic factors are undoubtedly relevant to meaning, it is difficult to account for them in corpus studies, where the language being analysed has been removed from its discourse context. It is therefore to the cotext that we must look.

Traditional studies of collocation typically start with a single word – the *node* – and note the patterns it forms with other words, in order to build up a picture of meaning through co-occurrence. Sinclair (1996) expands this practice with a view to identifying an *extended unit of meaning* by presenting an analysis which starts with the co-occurrence of word forms – collocation – and moves through progressively more abstract layers, addressing *colligation* (the co-occurrence of a grammatical class, such as possessive pronouns), *semantic preference* (collocation by semantic class rather than simply by word-form) and *semantic prosody* (a ‘consistent aura of meaning with which a form is imbued by its collocates’ Louw 1993:157). All of these elements of the extended unit of meaning will be called upon in the course of this paper, as they supply the features necessary for the disambiguation of sense – not only of single-word nodes, but also for multi-word units.

1.1 Collocational patternings and the establishment of canonical forms

If, as the psycholinguistic literature affirms (see especially Titone and Connine, 1999), multi-word units are learned as single lexical items, then they too should attract their own cotextual patterns in precisely the same way as single-word nodes do. There is no reason, therefore, to devise a specific approach to the analysis of multi-word units. What does need to be taken into consideration, however, is the possibility that not all the components of the multi-word unit are fixed. Well-established multi-word units such as idioms (*like a red rag to a bull*), proverbs (*red sky at night...*), and terminological expressions (*red/yellow card*) are cited in lexicographical descriptions by their most typical realisation, their *canonical form*. The identification of canonical forms is necessary in lexicography, but it does not come without complications. In the first place,

it is necessary to identify which elements are core to the phrase and which are part of an extended phraseology. For example, in an idiom such as *like a red rag to a bull*, is the canonical form the extended phrase *like a red rag to a bull*, or is it shortened to *a red rag to a bull*, or simply *red rag to a bull*? Additionally, should the American alternant *like a red flag before a bull* be classed as a separate idiom or simply as a regional variation? The varying presentations of the canonical form adopted in different dictionaries bears witness to the problem of identifying a clear cut-off point between core elements and extended unit of meaning.

The presence of a considerable number of truncated and ellipsed¹ idioms in corpus data does raise some doubts about the validity of canonical forms. Truncated forms typically involve a two- or three-word core collocation (such as *black sheep*, *red rag*, *pot* and *kettle*; see examples 1-3).

1. `Kevin was always supposed to be the black sheep and Richard the blue-eyed boy.
2. You shouldn't assume that mistakes don't occur within the genes and that they don't affect the body's function after you're born. But for some reason this is a red rag for some people.
3. FANS of John Major (yes, there are such people) say that Lady Thatcher's comments about him are `a bit rich". Have they, I wonder, ever heard the words `pot" and `kettle"?#

These forms do not contain all the significant collocates of the canonical phrase (*family*, *bull*, and *black* are all missing from the respective examples above). Yet despite the absence of these components, the meaning of the whole is understood, unless the expression is unfamiliar, in which case it is processed compositionally (Peleg *et al.* 2001). Although not all idioms appear in a truncated form, it is worth noting that in the case of the examples presented in this paper, and other truncated forms examined, the collocates that tend not to form part of the truncated form are those which are most likely to be exploited creatively in the coining of non-canonical forms: this will be discussed in 3.2 and 3.3.

The fact that truncated forms not only exist but are also quite common both in speech and writing suggests that each idiom has a core of salient lexical items which trigger the meaning of the conventional phrase, even in the absence of some of the other component parts of the canonical form. The less salient components can undergo variation provided that the central core is left intact. Conversely, when the salient core itself undergoes alteration, the cotextual patterning invariably corresponds to that which typically accompanies the full canonical form (Philip 2004). In other words, the salient core acts as the focus for variant realisations around it; when it is altered, the original, full form is recalled by the presence of conventional cotextual patternings². In Hoey's (2003) terminology, the core is *primed* for use with the other components of the canonical phrase: it is habitually associated with those other elements, and when they are absent they are filled in mentally in order to complete the remainder of the pre-learned sequence. Moon's notion of *idiom schemas* is also of relevance in explaining "the ease with which allusions to [fixed expressions or idioms] or exploitations are decoded" (1998:163), not only with regard to the idiom families that she describes, but also in the case of unconventional variants of multi-word expressions. These schemas are "characterized by an underlying conceit ... and an over-lying preferred lexical realization" (ibid: 163). The lexical realisation can be variable, so long as enough of the salient elements remain to tie the *underlying conceit* to the learned, conventional form,

thus allowing the listener to reconstruct those parts that are either missing entirely, or changed in form.

Although these observations do not help define what is and is not canonical, they do bring us closer to understanding what multi-word units are and how to identify them. The evidence points toward there being a salient core collocation which interacts closely with one or more semi-fixed collocates within a semi-fixed phraseology. The core collocation will normally correspond to the truncated version of the canonical form (which appears to contain the minimum number of lexical items necessary for the whole to be identified); this may, as was stated earlier, consist of more than two words, and they do not necessarily occur side by side (as illustrated in example 3).

1.2 Idiomaticity and delexicalisation

Traditional language reference works tend to be based on the underlying notion that language is organised by grammar elements which have a fixed rule-based order, and lexical elements, free of any constraint, which are inserted in the spaces in the grammatical structure. The phraseology of conventional multi-word units indicates rather that it is the grammar which is inserted to fill the gaps left by the lexical elements. It would certainly appear to be the case, given the considerations made above with regard to canonical and truncated forms, that once the salient core of a multi-word unit is selected, its immediate phraseology is more-or-less predetermined, and the context within which the extended phrase is located is, too, relatively predictable. This observation is in agreement with Sinclair's (1991) *idiom principle*, according to which "a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments." (Sinclair 1991:110). The idiom principle overrides its compositionally oriented counterpart, the *open-choice principle* (ibid.), in that the internal grammar of a phrase is established together with the lexical elements rather than as a sequence of independent choices. The resulting whole is interpreted as a single language choice, and although changes can occur in both grammar and lexis within the extended phrase, such variation is fairly restricted so as to conserve enough of the original meaning for the underlying phrase to be accessed (Philip 2000).

The idiom principle highlights the ubiquity with which ready-made language chunks of varying fixity are exploited (the *phraseological tendency*, Sinclair 1996: 82), shifting the focus away from one-word-at-a-time models of language (the *terminological tendency*, ibid.). A necessary correlate of phraseological selection is that meaning "is expressed, not by individual words but by phrases, and that it is phrases that can be said to have meaning, not individual words." (Hunston 2000:234). The blurring of word boundaries within phrases corresponds to a blurring of the semantic weight of each of those words which, operating in combination, lose some of their semantic distinctiveness. This effect is known as *semantic bleaching*, *semantic depletion* or *delexicalisation*³. In both idiomatic and delexical language, the combined meanings of the words are not equal to the overall meaning of the chunk, but the precise nature of the mismatch between compositional and non-compositional meaning differs. In proper idioms, the meaning of the whole is often palpably different from the compositional meaning, and even in the case of transparent idioms the meaning goes beyond the sum

of the parts to offer some additional sense that the a compositional analysis could not predict. For example, the transparent idiom *to give someone the green light* recalls the green traffic light (allowing someone to proceed), but the complete meaning of the expression is to authorise someone to begin a project or put plans into action. This extends and builds upon the compositional meaning of the words used.

In contrast, delexicalised strings generally mean something *less* than the sum of their parts. *Grey matter* is indeed (brain) matter which is greyish in colour; when used as a medical term, the meaning is compositional and contrasts with *white matter*, but outside medical terminology is used non-compositionally as a synonym for intelligence. The use of the expression does not imply that speakers are aware that the brain matter relating to intelligence is grey in colour, nor that the word *matter* refers to brain matter. In much the same way, Wray reports (2002:3-4) that many people are surprised to find out that the main ingredient of the breakfast cereal *Rice Krispies* is actually rice. The brand name has become distanced from the meaning of the component parts, resulting in the delexicalised meaning being more readily accessible than the compositional one. Delexicalised word strings are therefore not idiomatic as such, because they are no longer able to exert their full meaning potential, let alone support secondary meanings. Yet both types of phraseological language operate along similar lines: the more familiar they are, the more likely they are to be processed holistically, and although it is possible for them to be processed compositionally, this tends not to happen. Thus a key feature of multi-word units, as opposed to stable collocations, is the extent to which they are co-selected as single lexical counters whose combined meaning is delexicalised and therefore different to their compositional meaning.

2. Procedures for retrieving non-canonical variations of fixed phrases in corpora

The advantage of using a corpus to study multi-word expressions is that it is possible to study not only the known, canonical forms, but also variant forms of these. Corpus linguists are on the whole pessimistic with regard to the retrieval of non-canonical forms. Deignan (1999b: 197) states that “the researcher will only uncover what he or she sets out to look for”, and Moon (1998: 51) describes the successful discovery of variant forms as “ultimately a matter of serendipity”. This section will set forth a method for trawling a large general reference corpus for an inclusive data set which can then undergo refinement during analysis. Even here, no claim of exhaustiveness can realistically be made, as there is no way of ascertaining whether or not every pertinent example present in the corpus has in fact been retrieved. However, there are two considerable advantages in approaching the study of multi-word units in this way rather than by relying on data which has been collected opportunistically. The first is that a corpus provides a homogeneous data set in which the patterning of canonical forms can be compared to that of the non-canonical variants created at the same point in time, within the same text types and genres. The second is that by analysing variants all along the continuum from canonical to innovative, the principal trends can be identified and subsequently abstracted out to posit hypothetical patterns in a larger data set, such as the World Wide Web.

2.1 General considerations

The first stage in looking for variations to the canonical form is to understand the structural patterning of the canonical version; for this reason it is necessary to start by retrieving all occurrences of the canonical form, starting with a simple core collocation such as *red rag* in the case of the idiom *like a red rag to a bull*, and *black sheep* in *the black sheep of the family*. The resulting concordance lines provide a basis upon which a full analysis can be carried out, grouping and classifying the concordance lines in terms of their collocational and colligational features, and on their *semantic preferences* (Sinclair 1996). The abstraction upwards from simple word collocation to the identification of the semantic preference – the co-occurrence of collocates belonging to a particular semantic class – is of particular importance at this stage, because a key characteristic of everyday variation of conventional phraseological expressions lies in the substitution of one word for another which belongs to the same semantic class (Moon 1998; Partington 1996; Philip 2004, 2000). The establishment of such semantic classes as collocationally significant, and the identification of recurrent grammatical structures which confirm or differ from the full canonical expression in question, lead to the identification of typical structures, consisting of a series of slots, including the node, which can be likened to Sinclair and Renouf's (1991) *collocational frameworks*.

Once the principal collocates of the canonical phrase have been identified, along with their positions before or after the node, the analyst can posit a number of typical schematic structures in which the node is found. These patternings are then rendered as queries within the retrieval software being used in such a way as to allow individual slots to be left open, thus allowing alternant forms to appear in the results. It also needs to be borne in mind that one slot does not necessarily correspond to one word, because the variants themselves may consist of multi-word elements, or indeed omit some component of the typical framework. For this reason, some leeway has to be built into the collocational frameworks to allow for syntactic flexibility. The most straightforward means of achieving this is through the use of a facility which makes it possible to specify the minimum/ maximum number of words occurring between search items. Specific examples of collocational frameworks will be given in the case studies below (2.2 and 2.3).

It is worth noting that because the frameworks are initially based on the canonical form, the less a variant deviates from this, the more likely it will be to reappear during a series of related searches. It goes without saying that all duplicates must be eliminated from the resulting concordance lines before proceeding with the analysis. A further eventuality to consider is that a number of variant forms might display similar features which do not appear in the canonical form. Clearly, if this were to happen, the frameworks could undergo modification or extension with a view to carrying out further retrieval of variants. The lines obtained as a result of the combined search procedures have to be checked to eliminate the rogue forms that inevitably appear alongside the sought-for variants. This typically happens when the string crosses a sentence or clause boundary, causing the words in the search string to lose their collocational pull; it is also fairly common for unrelated collocations to appear, depending on the particular syntax used for the search string (see the rogue example *Red Bull* in the *red rag* search procedure, below). However, it is advisable to include all the examples which are not obviously *unrelated*, because it is easier to eliminate these at a later stage than recover

capitalised word initials. The other principal collocation which occurred was *Red Bull*, a caffeinated sports drink, which does not appear to have any relevance to the idiom in question. The concordance lines for this collocation were subsequently eliminated.

The next step was to search for variations to *bull*, and this was done by entering the basic canonical collocations as search terms:

- v) red+rag [the string *red rag* with no words intervening]
- vi) red+flag [the string *red flag* with no words intervening]

Whereas *red rag* provided only a couple of examples to add to those found in the first four searches, *red flag* proved much more problematic. This collocation is used in contexts outside the idiom, with several contextually-defined meanings (the communist red flag; a warning sign; the literal use meaning a flag which is red). To resolve this problem, the search was refined to include a preposition slot after the collocation. The concordance lines for *red flag* were alphabetically sorted to the right of the node, and those not followed by a preposition were deselected by hand.

As a final check for very innovative variations, the following search string was entered to find alternatives to *bull*.

- vii) like+a+1,5to+a [the strings *like a* followed by *to a* with up to five words intervening]

Of the resulting concordance lines, none had not been previously retrieved following steps one to six; the data set was therefore considered complete. Duplicate examples were removed from the combined file containing the results of queries one to seven before analysis commenced.

2.3 Case study 2: *the black sheep of the family*

The black sheep of the family is more problematic an expression than *like a red rag to a bull* because the central collocation, *black sheep* is used not only in the idiomatic expression but also as a biblical metaphor (from which the idiom is derived), as well as in the literal sense to refer to sheep with a black or dark fleece. Whereas both the metaphor and the idiom are considered to be multi-word units, the literal meaning is simply a compound of the colour, black, and the animal, sheep. The task, then, was not only to retrieve all the occurrences of *black sheep* which relate to the idiomatic and metaphorical senses, as well as relevant variant forms, but also how to identify the literal use from these so that it could be excluded from the data set.

First of all, the central collocation was retrieved (search i), and the collocational analysis carried out. As one might expect of a reasonably common idiom such as this one, a fair number of concordance lines (195) were retrieved in this simple search, covering all the different meanings outlined above. The central node was then tested by carrying out searches ii and iii, both of which were based on the node's most frequent collocation, *family*:

- i) black+sheep [the string *black sheep* with no words intervening]
- ii) the+1,5of+the+family [*the* followed by the string *of the family* with up to five words intervening]
- iii) sheep+1,5family [*sheep* followed by *family* with up to five words intervening]

Search two gave rise to a host of expressions which did not relate to the idiom, but which located *black sheep* within a paradigm of expressions that describe family members in terms of kinship (*daughter, mother, son, uncle*, etc.) and subjective attributes (*baby, brains, practical joker*, etc.). These do not add to the data set required to study the idiom, although they do highlight that the idiom itself is an institutionalised variant of a more general structure that is used to describe members of family groups, based on a simple genitive framework. Search two also provided four exploitations of the idioms involving variation to the colour *black*. Search three provided no relevant concordance lines that had not been retrieved in search one.

In order to formulate further search procedures, it was necessary to identify frameworks from the patterns that arose in the attested variation. Although the most significant collocate in the context of the multi-word node was *family*, it was often pre-modified, and therefore had no easily determinable syntactic position relative to the node. It could also appear in its possessive form before the node, which in turn altered the syntactic make-up of the idiom as a whole from *the black sheep of the family* to *the family's black sheep*. Furthermore, the expression *in the family* was quite common as a variant of *of the family*. A further series of searches was carried out in order to ascertain the frequency with which alternative colour words could be used, in addition to those discovered in search ii. These followed the pattern of the central collocation adopted in search one, substituting *black* for other colour words. This gave rise to no further exploitations of the idiom. The concordances found for *black sheep* are shown in Appendix 1.

2.4 Commentary

The procedures outlined in 2.1, and illustrated in 2.2 and 2.3, serve merely to compile the data set but do not in themselves constitute an analysis. By reducing the canonical form of the multi-word unit to its smallest realisation, a considerable number of examples can be collected, but not all of these will be directly relevant to the expression under study. Some of these are rogue forms which can be eliminated outright (for example *Red Bull* in 2.2), but the vast majority will be retained, including the literal, compositional examples. It is important to consider the compositional realisations of the word string in question because there is substantial evidence to suggest that literal and metaphorical meanings of words - and phrases - attract different and distinct collocational patternings (Deignan 1999a, 1999b). Therefore, one of the ways in which the multi-word unit will be distinguished from its compositional counterpart will be in the different collocates found in the surrounding context.

Although the procedures outlined above seem cumbersome, they privilege the retrieval of relevant data rather than the analyst's preconceived notions of what the data should be. Other corpus studies dealing with variant forms of multi-word units tend to assume

that the key collocates will not themselves be varied, or will look for conventionalised variant forms by entering separate queries for members of the semantic set to which the key words belong (see Cicogni and Coffey 2000; Moon 1996, 1998). These assumptions regarding variation are of course valid, but they are not flexible enough to allow the full range of variation to appear. An extended set of search procedures, as outlined in 2.2 and 2.3, allows the multi-word unit to be grouped with lexically similar forms, whether or not they are semantically related. This results in an analysis which is concerned both with differentiating the compositional string from the non-compositional string, and with accounting for the types of variation that can occur, the textual effect that these variations have, and the parameters that govern the identification of the multi-word unit when it appears in non-canonical form.

3. Identification and classification of multi-word units

3.1 Compositional and non-compositional collocations

We saw in 2.1-2.3 that one of the most effective and reliable means of identifying a multi-word unit is by treating a core, salient collocation as its base form. This collocation invariably corresponds to the truncated form found in both spoken and written forms of the language. However, this core collocation need not be exclusively attached to the multi-word unit in question: it may simply be a fixed collocation, such as *black sheep* or *grey matter* (examples 4-5); it may be a compositional collocation such as *red rag* (example 6); it may allude to the multi-word unit, but yet not constitute an example of it, such as the stage-name of a rapper, *Black Sheep*, and the name of the feminist-socialist newspaper *Red Rag* (examples 7-8). For this reason, the cotextual environment of the collocation needs to be analysed in order to home in on the multi-word unit. Although this can be done by hand with no great difficulty, there are certain considerations that must be made if one is to extrapolate automatic retrieval and classification procedures from the manual ones.

4. <MOX> Why do white sheep eat more than black sheep? <FOX> Yes because <ZGY>
<MOX> Sounds like a joke.
5. Albert Einstein agreed to have his brain removed after death, knowing that the grey matter that discovered the theory of relativity would attract scientific curiosity.
6. That scraggy bit of red rag knotted around his plump, whiskery-grey neck.
7. They joined forces with rap stars Black Sheep, Grand Puba, Gang Starr and Main Source for their CD
8. In 1978 Bea Campbell and Val Charlton wrote 'Work to rule' in Red Rag in a mood of despair over the debacle of Labour's economic policies.

In the first place, although the canonical form can be extracted straight away because its lexico-syntactic structure is already known, this form accounts for relatively few of the total number of examples pertaining to the unit as a whole. How should the remaining examples be treated? One method would of course be to ignore all other forms, be they variants or simply truncated forms, even though this is no real solution in that it only postpones the time when the problem of non-canonical forms will have to be addressed. From the examples supplied above, it has been shown quite conclusively that the canonical form is not necessarily the only acceptable form of a multi-word unit, therefore the alternatives must be taken into consideration to some extent. On the other hand, deliberate variation to the core collocation, of the type found in punning and other

word-play, can be set to one side for several reasons. The first of these is that this type of variation is very rare: in a search for variant colour words in the expression *catch red handed*, only eleven examples were retrieved by Google, despite its considerable size (Philip 2004). A second point in favour of their exclusion is that they are limited to very specific types of language (mainly tabloid journalism and advertising) and do not occur in the spoken data examined. Finally, variation of this sort is interesting semantically and on a text-linguistics level, but its study provides little if any insight into the nature of the multi-word units upon which the variation is based because the examples examined show that the cotextual patterning does not deviate from the norm identified for the full canonical form. This is almost certainly due to the fact that the pragmatics of communication tend to prevent people from being overly creative with language, as the result can all too often lead to communicative failure.

3.2 Truncated and ellipsed forms

Truncated and ellipsed forms come in various guises. Missing salient collocates often appear in the extended cotext – a window of ten words before and after the node is often sufficient to find these; but this is only true of those examples which otherwise behave in the same way as the full citational form, such as examples 9 and 10.

9. Pieret was a talented black sheep - the son of a respectable Brussels attorney, who might have succeeded in a variety of professions if he had not derived so much pleasure from being a psychopath.

10. I hear you're doing a history of the Sandon family, Mr Selby," Donald Bakewell remarked. 'Lot of black sheep there, I imagine.

Other types of truncation typically involve names and titles which may allude to the non-compositional multi-word unit without exploiting it productively (as is the case with *Red Rag*, example 8), but equally, they may refer to the compositional meaning of the collocate in question. In some cases, especially titles and headings, there is deliberate word-play at work which involves the processing of both the compositional and the non-compositional meaning simultaneously. Although titles attract cotextual features which mark them out as such (capitalised word initials being top amongst these), this is not sufficient in itself to decipher the meaning, which is typically unravelled in the course of the text that follows. This is clearly less problematic for newspaper headlines and the like, which precede relatively short texts, than it is for film, book, or magazine titles (examples 11, 12, and 8 respectively). Here the connection is more elusive, and although the cotext of the example may hint at the meaning intended, as in examples 8 and 9, knowledge of the content of the script or text may be fundamental to its correct interpretation.

11. <p> Black Sheep <p> David Spade volunteers to baby-sit Mike. (Chris Farley) a lovable accident waiting to happen. Big mistake. Rated: PG – 13.

12. Italo Calvino wrote *The Black Sheep* in 1943, when he was only 20, and yet as an Italian political allegory its relevance is untarnished.

Truncated forms which are not capitalised are more difficult to distinguish from the compositional collocation than those which are, but the feature that marks out the compositional collocate is its grounding in a literal context. Literally-grounding

collocates are those which confirm a literal interpretation by virtue of their relationship to the node in an extra-linguistic context, such as *bleat, farm, graze* and *wool* in the case of *black sheep*. The presence of such collocates favours the literal meaning by making it more contextually salient (Peleg *et al.* 2001) than the non-compositional one. The more of these collocates there are in the context, the more solid the literal grounding will be, as demonstrated in examples 13 and 14.

13. But Menaker doesn't buy the fur defence. `Fur is trivial," he says. `In a black sheep it might not be trivial but in a white rat it's trivial.

14. The two wool-spinning sisters, thought to be named Bayliss, who gave Weavers its name, moved into the house after the Second World War. They took to weaving the wool from the black sheep kept on a local farm, after realising nobody else wanted it.

Truncated forms of multi-word units offer fewer easily-identifiable features to aid their correct interpretation. This is especially true when the non-compositional form has more than one possible meaning, as is the case for *black sheep*, which has a metaphorical sense originating in the Bible from which the idiomatic *black sheep of the family* is derived. Similarly, the metaphorical use is difficult to distinguish from the literal use as both meanings ostensibly share a number of collocates (*flock, fold, shepherd, pastor*). However Deignan (1999a, 1999b) has demonstrated that collocational patternings and context play a determining role in ascertaining whether a word is literal or otherwise. Interestingly, she notes that the collocational pull operating between some words is so highly restricted that the collocational tendency cannot be extended to other members of the same semantic groups. As she explains:

It seems possible that once particular collocations become associated with a particular sense of a word, speakers are reluctant to use them with a different sense of that word. To do so might lead hearers to process a different sense from the one intended. Some collocations may become associated with a target domain sense of their component words, and speakers then avoid using these collocations in the source domain....Thus the tendency to map creatively and intellectually from source to target domains is restrained by a conflicting tendency – to fix and reuse conventionalized strings. (Deignan 1999a: 34)

So although it is quite possible to use a collocate associated with one particular meaning of a word in combination with another meaning of the same word, speakers tend not to do so because they possess some level of awareness that the consequence will be a loss of communicative effectiveness. To return to *black sheep*, analysis of the data shows that *pastor* does not occur at all with the literal, compositional meaning, and *shepherd* does not occur with the metaphorical sense. *Flock* and *fold* are problematic because they have both a literal and a metaphorical meaning, although in the data examined, *fold* never appears with the compositional collocation. The differentiation of meanings here relies not on some internal feature of the core collocation, nor on the identification in the co-text of a collocate to make one interpretation more salient than another, but on the interpretation of the collocate itself as being literal or metaphorical. This identification process is not a chance event. The collocates are in turn confirmed as being literal or metaphorical by their own collocates, such as *family, fraternity* and, in example 16, *pastor* and *dog-collar* (this latter making clear that the *pastor* is a clergyman by inserting a variant element into the conventional phrase *to be up to one's neck in something*). Compare the collocates in examples 15 and 16 with those in 17.

15. Christmas is that mistily sentimental time of family togetherness. Black sheep are welcomed back into the flock, and made to wear hideous Aran sweaters knitted by Auntie Pat.

16. Alex, pastor of the largest flock of black sheep in London, agreed that he was up to his dog collar in Easter weddings and it was doing his head in.

17. Whenever one of the white sheep bleats a black sheep crosses the river and joins the white flock, turning white itself in the process. When one of the black sheep bleats, the opposite occurs.

The evidence suggests that there is a kind of collocational domino-effect in operation, whereby literal or metaphorical meanings are determined by their collocates, which are subsequently defined as literal or metaphorical by their collocates. This has further implications for the retrieval of non-compositional strings because the contextual cues which aid in their interpretation will no longer be restricted to the unit itself, nor to its immediate collocates, but to its secondary, even tertiary, collocates (those terms which collocate with the unit's collocates). This domino effect in collocation supplies the ill-defined "contextual and pragmatic cues" (Wray 2002:31) that the human analyst searches for manually by examining the extended context in order to establish which meaning is intended. Though the analyst may not be able to pinpoint those features which prepare the ground for a literal, compositional interpretation, or a non-compositional one, they are undoubtedly present. It is quite possible that the reason why these features are difficult to quantify is because they relate not to the immediate collocates being examined, but rather to collocates further down the chain somewhere in the extended context.

3.3 Non-canonical forms and variation by semantic class

Instances of non-canonical multi-word units which involve the semantic extension of a collocate are difficult to categorise, but in a different way from truncated forms. Earlier in this paper it was shown that canonical forms are difficult to define with any accuracy, and that it is probably more useful to talk of core collocations which attract a particularly stable collocational environment. In fact, citational canonical forms of the type found in language reference works are simultaneously over- and under-specific: they over-specify the extent of the fixed core collocation, to include the variable non-core collocates, and they under-specify the variation that typically occurs to these non-core collocates. The variation that can occur to salient collocates of the core of a multi-word unit creates problems for analysts and machines alike because it appears to be infinitely productive and devoid of ground-rules governing what is acceptable and what is not, because although the novel element always relates to the term it supplants, this relationship may not be conventional, and may not follow the traditional rules of semantic class membership. Consider the following examples (see also Appendix 2):

18. The Marquess of Bristol, black sheep of the nobility, died on January 10th, aged 44

19. The 29-year-old Mercatone rider said that since the drugs scandal at last year's Tour de France, cyclists had become the black sheep of the sporting world.

20. China's leader arrived in Japan just as the black sheep of the country's ruling party returned to form a new coalition.

21. Indeed, Peter Rabbit in his own miniscule way is one of the true black sheep of literature.

22. THE life assurance industry - for long the black sheep of Britain's financial services - is

struggling to mend its ways.

23. Barnaby Gaitlin is a loser, just short of 30. He's the black sheep of a philanthropic Baltimore family.

24. THE black sheep of one of Britain's wealthiest families, escaped jail yesterday despite admitting possessing cocaine and crack cocaine

It is with no great difficulty that the human analyst can classify examples 18-22 as all belonging to the same category as the canonical collocate, *family*; a machine would find the leap far more problematic.

Glucksberg and McGlone's *class inclusion hypothesis* (1999) is of particular relevance in explaining the mechanisms underlying this type of variation. Although semantic classes are traditionally considered to be taxonomic and based on literal interpretations of a superordinate, Glucksberg and McGlone's view is that they tend to be *attributive* in non-literal and metaphorical language. The attributes which unite the variant forms of the salient collocation in examples 18-22 are shared with *family*: they can all be construed of as more-or-less homogeneous groups comprising human members. However the same expressions might not trigger any association with *family* in a different context. The core collocation of the multi-word unit typically predicts the conventional collocation *family*, but when this is substituted, the listener makes sense of it by finding qualities in the new element which are shared by the collocation which is already known.

According to our attributive categorization view, a metaphor vehicle, in the context of a specific metaphor topic, acts as a cue for the speaker to infer or construct a relevant category to which both vehicle and topic belong, with the following important constraints. The metaphor vehicle must, to some degree, epitomize or symbolize that category. The metaphor topic, by virtue of being assigned to that category, is characterized along one or more relevant dimensions.

(Glucksberg and McGlone 1999: 1546)

The likelihood of there being a collocate that cannot be classified alongside the other members of the *family* paradigm is very low because the presence of a variant element simultaneously triggers the interpretation of the conventional form together with the new. As the conventional item here is *family*, the class-inclusion hypothesis indicates that any variant would be expected to share some attributes of that class. This occurs because whereas a taxonomic category has a finite set of characteristics and members, an attributive category is much more open-ended, allowing the hearer a considerable degree of interpretative flexibility. In a variant form of a multi-word unit, the similarity between the new element and that which it replaces is not absolute, and so does not rely on the traditional notion of semantic class; rather it operates at a more selective level, highlighting or selecting particular characteristics according to their appropriateness. The more distant the relationship between the novel and conventional elements, the more general the shared attributes are, but a connection is made nonetheless by virtue of the expectations set up by prior knowledge of the conventional collocate.

As a means of testing for possible sources of shared attributes, a series of WordNet searches was carried out for *black sheep*. The hypernym search placed the attributes relating to "person, individual, someone, somebody, mortal, human, soul" at the fourth level in the hierarchy in sense 1 ("a reckless and unprincipled reprobate"). This is the

most general category in the hierarchy before “organism, being” which is shared by both the literal, compositional and the non-compositional senses of *black sheep*. Given this knowledge, it can be expected that collocates pertaining to the human world are all theoretically possible when the sense is non-compositional, and those collocates relating to the animal kingdom will be likely to have a literally-grounding effect on the meaning, because the next level up in the hierarchy is too general to allow a sense distinction to be made reliably. So the fact that the groups in examples 18-22 involve humans is key to the interpretation of *black sheep* as belonging to the multi-word unit rather than to the compositional phrase; the compositional interpretation, as was seen in examples 13 and 14 above, is triggered by collocates such as *farm, cow, white sheep, wool* and *fur*, which all relate to animals.

The difference between the traditional notion of semantic class and the groups appearing in the examples offered here are satisfactorily explained by Glucksberg and McGlone’s class inclusion hypothesis. Although the traditional *black sheep* is one of the *family*, it can just as well belong to a non-standard family group, such as *the lending sector, the sporting world* or *Thai Buddhism*. A metaphorical *family* need not involve kinship, but does necessitate some form of group identity, and it is this attribute that drives the creative substitution in *black sheep*. Similarly, Philip (2000: 226-7) discussed the effects of substitution in *like a red rag to a bull*, observing that instead of *bull*, variants such as *the purists, the true blues, the Americans, the Euro-sceptics*, and so on, can be found. The attributes traditionally associated with *bulls* are different to those associated with *family*, and for this reason the substitutes for *bull* are united by characteristics such as stolidity, stubbornness, resistance to change, anger and strength. However, being human, the substituted entities cannot be subsumed under a general heading such as “male bovids”.

It can be noticed that the canonical schema is preserved despite variation: there is no evidence of anyone being described as the *black sheep* of anything other than a group or community, and no evidence of a *red rag* provoking anything other than bulls and particular classes of stubborn and opinionated humans. In linguistic terms, this means that the semantic preferences ascertained in the study of canonical examples involving the key-words *red rag* and *black sheep* are adhered to despite variation, although it must be stressed that semantic preference, in these cases at least, is based on attributive categorisation, and not on the traditional, lexicographic notion of semantic class. From a computational point of view, this makes a definition of the variable collocate slot more difficult. When a semantic set is lexicographic, a thesaurus is all one needs in order to access the paradigm of lexical choice. When the set is constructed around shared attributes – which are as likely to be metaphorical or connotative as they are to be literal – the matter is substantially more complicated. On the other hand, computational knowledge bases can gather information from non-canonical forms thanks to the fact that the canonical phraseology predicts a variable slot. While the conventional members of this slot will identify themselves as being the most frequent collocates, the infrequent unconventional ones serve to pad out a category with shared attributes which might be impossible to quantify on the basis of traditional semantic sets. This means that attested variation may contribute to the construction of more sophisticated semantic knowledge which could subsequently find application in identification, extraction and classification.

3.4 Substitution by semantic class and the extended unit of meaning: semantic preference and semantic prosody

Although substitution by semantic class is a very wide-ranging type of variation, it can be reduced to two generic types – the *incidental* and the *intentional*. Incidental substitution involves near-synonymy and, as Moon points out, “there is no real change in meaning of the [fixed expression or idiom], although there may be register distinctions” (1998: 124). In contrast, intentional substitution picks up on elements of the extended multi-word unit’s phraseology in order to emphasise them in some way. Whereas incidental substitution tends to occur with words of the same rank, intentional substitution is often an emphatic and focusing device based upon the use of hyponyms.

Taking *into the red* as an example, the incidental substitution observed occurs by replacing the verb *go* with another verb of movement such as *move*, *run* or *head*. The intentional substitution, on the other hand, involves *dive*, *plunge*, *slip*, *slide*, and so on, which are still verbs of movement, though they differ on two counts from those involved in incidental substitution. In the first place, they re-focus the direction of movement from advancing to falling, and in so doing they invoke and place emphasis on the institutionalised metaphor in which numerical reduction is expressed as downward movement. By emphasising the downward movement, the delexicalised, non-compositional phrase undergoes a degree of re-lexicalisation which shifts the interpretation a step back in the direction of compositionality. The second difference is that each of these intentionally-substituted verbs has its own semantic focus and connotative value within the broader concept of downwards movement.

Substitution of this type can be said to operate within the framework of *semantic preference* (Sinclair 1996) rather than the more general concept of semantic sets or near-synonyms. For example, *cuss*, *swear*, and *scream* are all present as substitutions for *talk* in the context of *[talk] a blue streak*, though it may be noted that other verbs of discourse such as *discuss* and *chat*, ostensibly belonging to the same semantic set, are not present. The substituted verbs conform to the semantic class of the original but add an emphatic tinge which corresponds to and highlights the semantics of the remainder of the phrase, which is re-lexicalised to some extent as a result of the substitution. Verbs which relate to reasonable, polite discussion would be inappropriate in the context of this expression and therefore do not occur. In the same vein, all sorts of vocal activities can be carried out until one is *blue in the face* (see Appendix 3), including *answer*, *argue*, *ask*, *call*, *cite*, *criticise*, *deny*, *discuss*, *extol the virtues of...*, *lecture*, *offer advice*, *shout*, *speak*, *talk* and *tut-tut*. These are all verbs of verbal communication which, however, carry with them the implication of non-communication which is also present in the remainder of the phrase (...until you are *blue in the face*); in this way a semantic prosody is strongly associated with the semantic preference. Those verbs which are likely contenders for the slot are verbs of communication, but within this wide range of possible candidates, some are deselected as being inappropriate because they do not carry with them the semantic prosody of communicative difficulty. It may be interesting to note that the verbs *speak* and *talk*, which are often cited as forming part of the canonical phrase, appear to have no discernible semantic prosody. Their prosodic neutrality is likely to be connected to the fact that they are super-ordinates of the class

of speech verbs, which makes them semantically all-purpose and less likely to attract an identifiable semantic prosody than their hyponyms are. More specific verbs, such as *extol the virtues of...*, *offer advice* and *tut-tut*, attract a semantic prosody which reinforces and intensifies the meaning of the multi-word unit, and in so doing they re-introduce an element of compositionality into its overall processing.

Substitution makes it possible for the analyst to hone the definition of the non-core salient collocates. The canonical forms offer one or two terms which at first sight appear to undergo substitution by semantic class. Yet attested variation demonstrates that the class to which the variants belong is not a traditional semantic set but a more specific, restricted paradigm of choice whose function is quite clearly to reinforce the non-compositional meaning of the multi-word unit. Although variants may belong to the same lexicographic semantic set, as is true for *blue in the face*, those members of that set which are used are united by their shared attributes (in this example, the attributive class can be glossed as *vocal behaviour which falls upon deaf ears*). Members of the same semantic set which do not share these attributes are ineligible for the slot, except under special circumstances – most notably humour and irony.

4. Concluding remarks

Unlike most investigations into idiomatic and other non-compositional language, this paper has focused on identical strings which can be either compositional or non-compositional, depending on the cotext in which they occur, such as *in black and white*, *grey matter*, and *black sheep*. Well-known idioms such as *raining cats and dogs* or *kick the bucket* are easy to identify because they are learned as single, non-compositional chunks much in the same way as single words are learned (Titone and Connine 1999). There is no conflict between the compositional and non-compositional interpretations when it is the non-compositional meaning alone that is conventional and salient. It is less easy to decide which interpretation is required when more than one meaning of a multi-word unit is conventional.

Context of course plays the determining role in sense disambiguation. This paper illustrates some of the ways in which cotextual features support and define the meaning of a salient core collocation by building up extended units of meaning around it, in precisely the same way as such units form around single words. In treating the multi-word units in their most concise form – that which generally corresponds to the truncated or ellipsed form found in the corpus data – rather than the traditional, lexicographically-defined canonical form, it is possible to demonstrate how variations and incomplete citations are interpreted as belonging to the non-compositional unit rather than to the compositional collocation, and vice versa. Each meaning attracts different groups of collocates, so by taking account of the cotext surrounding the node, it becomes apparent which interpretation is relevant in a given context (see examples 13 and 14). The major exception to this is when the core collocation is used to exploit the cultural associations of the non-compositional phrase as a whole, as seen in examples 7 and 8.

So what is a multi-word unit? It would appear to be a fixed, non-compositional collocation which attracts semi-fixed salient collocates with shared attributes. Once the

unit has been learned, the fixed collocation is then primed to co-occur with the semi-fixed collocates(s). If the collocates(s) are creatively exploited, a connection is made between the variant and its conventional counterpart which is based on the identification of shared attributes: if no connection can be made, the semantic preference is violated. Should the collocates(s) be omitted entirely from its habitual position in the phraseology, as occurs in truncated and ellipsed forms, it has been observed that these are compensated for by the presence of related collocates elsewhere in the immediate context. It is therefore apparent that the encyclopaedic knowledge of a multi-word unit involves the identification of a canonical form of sorts, but this form is more flexible than lexicographic descriptions typically admit, involving a more schematic depiction of the syntactic position and the semantic features of the salient collocates.

In many ways, a multi-word unit is no different to a single word. It combines with collocates which refine its meaning, and with colligates which supply it with typical syntactic patterns. Part of its form is its collocability with a particular semantic group, one or two members of which tend to be considered canonical or standard; and within these variable collocates a semantic prosody can often be identified. On the other hand, multi-word units are different from orthographically-identical, compositional collocations, as this paper has shown. This difference lies in determining whether or not a collocation has begun to fossilize, resulting in a delexicalised, non-compositional chunk whose collocational environment is distinct from that of the compositional collocation.

Notes

1 Truncated forms retain the phraseology of the full form, but cut it off midway. Truncated forms are especially common in the citation of proverbs, for example *It's the last straw... [...that breaks the camel's back]*. Ellipsed forms pick up on salient elements of the full form but do not retain the original phraseology, for example in this corpus citation based on *the pot calling the kettle black*: "Call me picky but when the Exchequer makes eight quid from every tenner spent on petrol the phrase, 'Pot, kettle and black,' springs to mind". For the sake of clarity, these two phenomena will be treated under the single term *truncated forms* as the majority of the examples brought forward in this paper are truncated rather than ellipsed.

2 The data suggests that simultaneous variation of salient core and context is extremely rare (Philip 2003). When it is found, it tends to be restricted to transcribed spontaneous speech – which has not undergone the editing that written texts do – and seems to occur when two similar structures are confused during on-line processing.

3 *Delexicalisation* is the preferred term in this paper, relating as it does not only to semantic bleaching but also to a parallel phenomenon resulting in the depletion of distinct grammatical function.

References

- Cicogni, L & S. Coffey (2000). A Corpus study of Italian Proverbs: implications for lexicographical description. *Euralex 2000 proceedings* (pp. 549-555). Stuttgart: Universität Stuttgart.
- Deignan, A. (1999a). Linguistic Metaphors and Collocation in Nonliterary Corpus Data. *Metaphor and Symbol* 14 (1), 19-36
- Deignan, A. (1999b). Corpus-based research into metaphor. In L. Cameron & G. Low (Eds.), *Researching and Applying Metaphor* (pp. 177-199). Cambridge: Cambridge University Press.

- Gibbs, R.W. (2002). A new look at literal meaning in understanding what is said and implicated. *Journal of Pragmatics* 34, 457-486.
- Glucksberg, S & M.S. McGlone (1999). When love is not a journey: What metaphors mean. *Journal of Pragmatics* 31, 1541-1558.
- Hoey, M. (2003). Lexical Priming and the Properties of Text
<www.monabaker.com/tsresources/LexicalPrimingandthePropertiesofText.htm>
accessed 03/06/2003.
- Hunston, S. (2000). Phraseology and the modal verb: a study of pattern and meaning. In Heffer & Sauntson (Eds.) (pp. 234-248). *Words in Context: A Tribute to John Sinclair on his Retirement*. ELR Monograph 18. Birmingham: The University of Birmingham.
- Louw, B. (1993). Irony in the Text or Insincerity in the Writer?: The Diagnostic Potential of Semantic Prosodies. In Baker, M., G. Francis & E. Tognini-Bonelli. (Eds.) (pp. 157-176). *Text and Technology: In Honour of John Sinclair*. Philadelphia & Amsterdam: John Benjamin.
- Moon, R. (1998). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford: Clarendon.
- Moon, R. (1996). Data, Description, and Idioms in Corpus Lexicography. *Euralex '96 Proceedings*. (pp. 245-256). Gothenburg: Gothenburg University.
- Partington, A. (1996). *Patterns and Meanings: Using Corpora for English Language Research and Teaching*. Amsterdam and Philadelphia: John Benjamin.
- Peleg, O., R. Giora and O. Fein (2001). Salience and Context Effects: Two Are Better Than One. *Metaphor and Symbol* 16 (3-4), 173-192.
- Philip, G. (2004). Habeas Corpus: Direct Access, Salience and Delexicalisation in Corpus-based Metaphor Studies. Paper presented at *Mind, Language and Metaphor: EURESCO Conference on the Processing of Metaphor and Metonymy - From Computers to Neuropsychology*. Granada, Spain 24-29 April 2004.
- Philip, G. (2003). *Connotation and Collocation: A Corpus-Based Investigation of Colour Words in English and Italian*. PhD Thesis. Birmingham: The University of Birmingham.
- Philip, G. (2000). An Idiomatic Theme and Variations. In C. Heffer and H. Sauntson (Eds.) (pp. 221-233). *Words in Context: A Tribute to John Sinclair on his Retirement*. ELR Monograph 18. Birmingham: The University of Birmingham.
- Sinclair, J.M. (1996). The Search for Units of Meaning. *TEXTUS*, 9 (1), 71-106.
- Sinclair, J.M. (1991). *Corpus, Concordance, Collocation*. Oxford: OUP.
- Sinclair, J.M. and A. Renouf (1991). Collocational Frameworks in English. Reprinted in J.A. Foley (Ed.) (1996) (pp. 55-71) *J.M. Sinclair on Lexis and Lexicography*. Singapore: Unipress.
- Titone, D.A. & C.M. Connine (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics* 31, 1655-1674.
- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

APPENDIX 1: *black sheep*

1. indy/UK s concerned," ran a review of The White Black Sheep (1926). <p> Miller's Hollywood career lasted for 70
2. indy/UK itter led a revival in Geordie brewing. Black Sheep 1992. Masham, Yorkshire. A classic new brewery. Its
3. brmags/UK he furnaces of the jewellery family's black sheep. </p> <p> NOVEMBER 11 Christie's: Impressionist and
4. brspok/UK M01> Why do white sheep eat more than black sheep? <FOX> What? <MOX> <tc text=laughs> <FOX> Because th
5. brspok/UK MOX> Why do white sheep eat more than black sheep? <FOX> Yes because <ZGY> <MOX> Sounds like a joke.
6. brmags/UK Catholic family. Were you always the black sheep # <LTH> B.G. I was the pink sheep. My family weren't
7. brbooks/UKild. He reminded me, sweetly, <f> The Black Sheep <f> <p> <f> of the rewards and what, alas, awaited m
8. today/UK nity to rub shoulders with the Queen's black sheep. <p> For just $ 25,000 (around £ 14,500
9. today/UK to remind him that he's the Family's black sheep. <p> They cheered as Johnson, competitor number 226,
10. usbooks/US people.62 <p> A Sacred Cow Becomes a Black Sheep <p> The Highway Trust Fund, regarded in its early ye
11. npr/US ack Sheep one. <p> Excerpt of music from Black Sheep) <p> Hansen: Sounds like you had a lot of fun. Kinca
12. usephem/US comedy. Rated: PG <!--details--> <p> Black Sheep <p> David Spade volunteers to baby-sit Mike. (Chris
13. brbooks/UK ectic life and been a bit of a <151> black sheep; a gambler, and divorcee, but he was only in his
14. brmags/UKiam, is earning a name as a flamboyant black sheep. After battling over the family oil conglomerate, he
15. guard/UK Yeah, I get that," says Lou. `I'm the black sheep. All my sisters and brothers are teachers." You chos
16. brmags/UK </c> Roger Banks-Pye seems to by the black sheep amongst the Colefax and Fowler decorators. `I'm
17. times/UK </xr> <hl> Mutual friends not counting black sheep; Analysis </hl> <bl> Susan Emmett </bl> <dt> 21 May
18. usnews/USself-educated boys from poor families, black sheep and alcoholics," writes Ms. Milton. Crane, for exam
19. sunnow/UK: `Kevin was always supposed to be the black sheep and Richard the blue-eyed boy. In his mother's eyes,
20. npr/US rs: People have been painting this guy a black sheep and this guy that most people that know--knew him he
21. brbooks/UK genuinely communal effort run by the black sheep and not by the same old faces, forever looking for
22. brbooks/UKd, cherished and loudly acclaimed the Black Sheep and the Wicked Fairy, who carry the story forward? <
23. indy/UK /date> FEDERICO ZERI was at once doyen, black sheep and chief scourge of the Italian art-history world,
24. brmags/UK pgo choruses not. </p> <p> Masta Ace, Black Sheep and Ed OG are among the other wordsmiths doing it,
25. times/UK ntimental time of family togetherness. Black sheep are welcomed back into the flock, and made to wear
26. usbooks/USmple, `long-suffering # family star # black sheep") are very useful in identifying central, peripheral
27. indy/UK doggedness with which Rome's spiritual black sheep are being pursued by an administration that is left-
28. brephem/UKy are arts… <!--picture--> <h> Black Sheep Art Gallery </h> <!--address--> <!--telephone--> <p>
29. times/UK the night with my late friend, Nalin: black sheep, as he told me, not without pride, of a famous famil
30. guard/UK ion team, `and maybe have seven of the black sheep as clients, and if you are lucky two of them will be
31. today/UK e explains. <p> I'm the first one, the black sheep. At the beginning, it was difficult for my family to
32. sunnow/UKd bright." Christopher was a `bit of a black sheep." At 14, he was sent to a detention centre for shoot
33. sunnow/UKhe church of the good shepherd and the black sheep. At Kinnitty and in many other places I have discove
34. brspok/UKte well and I suppose in a way I'm the black sheep at the moment <M01> <ZGY> <F01> cos it's always me
35. guard/UK he mother, cast aside the image of the black sheep - backfired because no one had heard of El Ghriba or
36. brspok/UK What? <M01> White sheep eat more than black sheep because there are more white sheep. <FOX> I don't ge
37. brbooks/UKlack arts, black magic, black market, black sheep # Besides, it's inaccurate to say anyone is black or
38. brbooks/UKtself in the process. When one of the black sheep bleats, the opposite occurs. With slight modificatio

39. oznews/OZg Arts Complex. <sect id=WEEKEND> <hd> Black sheep blossoms </hd> GROWING up in Townsville and attendin
40. usnews/US Basie), and the deep, slow, sensuous "Black Sheep Blues" invite the players to swing. The writing is
41. indy/UK k ink into it and changed the title to "Black Sheep". Bridger gave his calling card to an assistant, who
42. sunnow/UKrce HQ in New Mexico and nicknamed the Black Sheep. British military experts said the markings suggeste
43. today/UK And it is refreshing to see Teddy, the black-sheep brother of John and Robert Kennedy, playing such a
44. usbooks/US by a lifelong resentment against his black-sheep brother, and soon the bride becomes the unwilling st
45. today/UKne." <p> And every royal family has its black sheep - but it is unclear whether LaToya will be pleased a
46. econ/UK once behaved as quasars; they were rare black sheep. But if the time for which a quasar is active is onl
47. econ/UK say that the railway company is zaito's black sheep"; but there seem to be plenty more in the flock. The
48. usbooks/USdown." <p> Cousin Arthur was a tricky black sheep, but he bore the name of Britain's hero-king. He doe
49. oznews/OZ idge, a yellow car called B rum and a black sheep called Kylemore all added up to a holiday unlike any
50. today/UK ennis </h> Graham Fisher <p> BLACK sheep Chris Wilkinson set up a second-round meeting with
51. today/UK Graham Fisher <p> BRITAIN'S black sheep Chris Wilkinson kept up his Wimbledon record with a
52. usbooks/USa" for "blackmail" and "outcast" for "black sheep." Clearly, some of these substitutions work better t
53. times/UK nectarines come from peaches, and how black sheep come from white -- the first steps towards recent
54. oznews/OZ ting Waits tunes in years: Hold On, a black-sheep cousin to the tunes Bruce Springsteen wrote for The
55. brbooks/UKnever one of the white sheep bleats a black sheep crosses the river and joins the white flock, turning
56. times/UK hy is more complex. His mother was the black sheep daughter of a wealthy shipping executive, a child of
57. today/UK l in EastEnders (BBC1, Mon, 8pm) - and black sheep David Wicks is her target. <p> Actress Michelle
58. brmags/UK as born into a family of eccentrics # Black sheep, each of them,' she says of her parents, and I was a
59. usephem/US rap stars! Main Source, Gang Starr, Black Sheep, etc. (Delicious Vinyl) E 1992 <!--catalogue-->
60. times/UK : "It's nothing to do with us. You get black sheep everywhere. It's none of my business if they weren't
61. brmags/UK Elvira-the camp vamp who looks like a black sheep form the Addam's Family. Since the demise of Lynx, P
62. npr/US ther. They joined forces with rap stars Black Sheep, Grand Puba, Gang Starr and Main Source for their CD
63. brbooks/UK upon two meadows divided by a river. Black sheep graze in one, white sheep in the other. Whenever one
64. today/UKscout, One Stab, but far from being the black sheep he is the one most loved by his father. <p> Growing
65. guard/UKwent into teaching, too. "But I was the black sheep," he says, not very convincingly, since he was head
66. indy/UK Ibiza with <KPD> 90 to his name. For a black sheep, he has been suprisingly productive. He has edited
67. brmags/UKm the sea gives him. "I was always the black sheep," he says. "Everyone else stayed in New Jersey. I wa
68. econ/UK on of the family. He was said to be the black sheep. His defenders say that at least he was a designer o
69. brbooks/UKher: "Whisky # <p> What <f> was <f> a black sheep, I asked myself. Obviously, in the general view, one
70. brbooks/UKer," he was saying to me. "The family black sheep. I'm rather proud of him, myself, but then, I've alw
71. times/UK into disarray. Italo Calvino wrote The Black Sheep in 1943, when he was only 20, and yet as an Italian
72. usbooks/US stories about her father's uncle, a black sheep in the family who had been an artist with an explosi
73. usbooks/US ing, but painful, label of being the black sheep in the family, always being teased about her "crazy
74. brbooks/UK blind us to the fact that there are black sheep in every fold and that a fraternity which started ou
75. brbooks/UKdigestive system. Caution} There are "black sheep" in the Compositae family, as all the artemisias, mo
76. brbooks/UK s, and a witch or two. There were no black sheep in Andersen - he would have found the idea
77. brbooks/UKit any other way.27 Charteris was the black sheep in Haig's family - the antithesis of the "gentlemen"
78. brbooks/UKse of me and my life. I was truly the black-sheep-in-the-closet-nightmare-relative-that-no-one-outside
79. guard/UKign, and concluded: "We do not want any black sheep in Thai Buddhism." But Dhammakaya, in Pathum Thani o

80. guard/UK). Alex, pastor of the largest flock of black sheep in London, agreed that he was up to his dog collar i
81. guard/UKreat as Cousin Bette, Old Goriot or The Black Sheep (in fact it's something of a potboiler) but there ar
82. indy/UK court action by his father. `There's a black sheep in all families," said the Duke, `There's nothing ne
83. brmags/UK al undyed black. Remember, we do have black sheep in the UK. </p><p> At the moment Esprit's fashion
84. sunnow/UK e from an elite squadron known as the Black Sheep-is designed to be `invisible" to radar. Experts had
85. newsi/UK ence. Fur is trivial," he says. `In a black sheep it might not be trivial but in a white rat it's
86. brbooks/UKa, every family in Greece has got its black sheep, it's been like that since the Civil War." <p> Leoni
87. guard/UK ne Gallery in London and re-titled it `Black Sheep". It is now back in the news for a different reason.
88. brbooks/UKots - and if he can, should he? Was a black sheep just a white sheep dirtied or black in his own right
89. times/UK They took to weaving the wool from the black sheep kept on a local farm, after realising nobody else
90. today/UKfamily member likely to cause trouble - black sheep La Toya Jackson - did not participate. <p> La Toya h
91. today/UKuperstar Michael, Janet and even family black sheep, LaToya. `The estate will be something where everyon
92. npr/US McIlhenny family rebelled? Is there some black sheep, maybe it would be a red sheep, in the family, who
93. sunnow/UK impression our mother was a bit of a black sheep. My grandfather had no photos to show us and didn't
94. brbooks/UK wanted to write about Bute-town, the black-sheep neighbourhood of my home town, and this looked like
95. guard/UKortion rallies and everything. Am I the black sheep? No, but they do pray for me all the time. I'm still
96. today/UKweek. <p> One ex-friend said: `He's the black sheep of the family now. People think we all have houses a
97. today/UKjust leaves the runt of the litter, the black sheep of the Douglas family, saddled with the decidedly un-
98. today/UKbarrel - and go for the WBO title, the black sheep of the world heavyweight championship belts, if he
99. today/UKd. <p> Grinning, he adds: `I'm just the black sheep of the family." <p> ASKED what his girlfriend thinks
100. today/UKister, who starred in Dallas, who's the black sheep of the family and I don't speak to. That certainly c
101. times/UKade, angel's trumpets and the notorious black sheep of the family -- tobacco. I am not talking about the
102. times/UKies most people's stereotype of a poet. Black sheep of a famous thespian family, ageless matinee looks,
103. times/UKN 0 099 27268 7. AN ODD-JOB man and the black sheep of a family that made its money from manufacturing
104. times/UK17 September 1999 </dt> John Amery, the black sheep of a distinguished family whose father was the
105. times/UKteiner </bl> <dt> 13 May 1999 </dt> THE black sheep of one of Britain's wealthiest families, escaped jai
106. times/UK security, looking upon herself as the `black sheep of the family". <p> I felt very rejected," she said
107. times/UKl, it was an opportunity to berate the `black sheep" of the family, the shame-faced Bradford & Bingl
108. times/UKuld still refer to him as `Charles, the black sheep of the family". He went to school at Oundle, where i
109. times/UK in the failed company. As a result the black sheep of the lending sector is now at the heart of the
110. usnews/USynex, which many analysts consider the black sheep of the Baby Bells, might be in line for another
111. econ/UK ss of Bristol: The Marquess of Bristol, black sheep of the nobility, died on January 10th, aged 44 </sub
112. econ/UKiss in his ancestry. He was not the sole black sheep of the family. One ancestor was an admiral who had a
113. econ/UKna's leader arrived in Japan just as the black sheep of the country's ruling party returned to form a new
114. oznews/OZr French captain Eric Cantona, another black sheep of international soccer. Our priority will be the
115. oznews/OZdna Gil in the title role. Malena, the black sheep of a noble family, takes many years to prove her own
116. oznews/OZis a loser, just short of 30. He's the black sheep of a philanthropic Baltimore family. He has an almos
117. oznews/OZine was banished from Melbourne as the black sheep of his blueblood family but new trainer Gerald Ryan
118. oznews/OZgton from all over Australia with `the black sheep of the family" travelling from New Zealand. `Yes, ev
119. oznews/OZte High School, Shannon Foley was `the black sheep of the family" who preferred theatre and dance to
120. oznews/OZold he was the illegitimate son of the black sheep of the family and a scarlet woman" in Sydney. Rule

121. oznews/OZour de France, cyclists had become the black sheep of the sporting world. <sect id=SPORT> <hd> TRIO MOT
122. usbooks/USunusual and anomalous `laborers", the black sheep of economic science: handmaidens and handymen, opera
123. usbooks/USis a man who deserted in the war, the black sheep of his family, but he's from a wealthy English
124. usbooks/USf what they are commanded to do. The `black sheep" of some families, young people who are heavy drinke
125. usbooks/USiards # At best, Spain is seen as the black sheep of the European family because, it is said, it did n
126. usbooks/USd no one. But it was his brother, the black sheep of the family, who beat his wife and children and di
127. usbooks/UShy was O'Keeffe considered to be the `black sheep of the family # How did the painter's family of
128. sunnow/UK urely will in the coming months. <hl> Black sheep of the Family; The Sun says; Leading article </hl> <
129. sunnow/UK her son. Christopher, though, was the black sheep of the family. He had been in and out of trouble fro
130. sunnow/UK Thornton -- and lied that he was the black-sheep of the famous chocolate family and had changed his n
131. sunnow/UK remier League level, but he'll be the black sheep of the family when shouting for United today. He adm
132. sunnow/UK led young Alex to be regarded as the black sheep of the family. For him, that period -- December 1963
133. sunnow/UK ses to comply, France will become the black sheep of the EU. News of the victory was flashed to Tony
134. sunnow/UK l hero John Aldridge took charge, the black sheep of the Merseyside family, has begun to fight back.
135. sunnow/UK ill manages to be the most unpleasant black sheep of the Mitchell family. Naturally he starts scroungi
136. brspok/UK ow I believe. So erm I'm a bit of the black sheep of the family there I'm afraid. I kept it a so quiet
137. brspok/UK k <ZGY> <ZZ1> <ZZ0> <ZGY> sort of the black sheep of the family anyway. <M01> <tc text=laughs>
138. brspok/UK . <F02> like so I was like one of the black sheep of the family. But my all my family except my two
139. brbooks/UKught, was a caddish sort of name. The black sheep of the family might be called it, the one despatched
140. brbooks/UKmachinery of the Children's Court the black sheep of the family can enjoy, in some public institution,
141. brbooks/UKng the matter for ever, `Esau was the black sheep of the family." Well, that was something I could
142. brbooks/UKre she lived, and because she was the black sheep of the family I was never encouraged to talk about h
143. brbooks/UK his own pad somewhere. If he was the black sheep of the family they were probably glad to have him go
144. brbooks/UKat Franco was resentful at seeing the black sheep of the family suddenly converted into a national her
145. brbooks/UK a bank. And poor Thaddeus became the black sheep of the family, according to Katie, wasting his
146. brbooks/UK own miniscule way is one of the true black sheep of literature. Like Alcott's Dan, he retains his
147. guard/UK tor" (the Action Man) and Duncan, (the black sheep) of the `X family" all seem to have a thing about
148. guard/UK ver any danger of Thomas ending up the black sheep of the family how could he outdo his paternal
149. guard/UK rfly? 12) Why is Charles de Gaulle the black sheep of the family? 13) If you log on to www. messiahcam.
150. guard/UK The move came as Yorkshire Water, the black sheep of the UK water industry during the 1995 drought whe
151. indy/UK r Gowen), is a sculptor, a self-hating black sheep of a Protestant ascendancy family and a repressed
152. indy/UK inal, as if they were the loud-mouthed black sheep of some fine and ancient family. Of course it goes
153. indy/UK life assurance industry - for long the black sheep of Britain's financial services - is struggling to m
154. indy/UK a pink jumpsuit. `I was definitely the black sheep of Queens! I got a lotta dirty looks. You were alway
155. indy/UK shocking reputation - 900 years as the black sheep of the marksmanship world, the wicked-upstart strain
156. brmags/UKow </p> <h> HEADLESS CHICKENS </h> <p> BLACK sheep of the label. </p> <p> Headless Chickens developed f
157. brmags/UKe rebellious daughter, a wanderer, the black sheep of the family." Jack Nicholson, who had earlier co-
158. brmags/UKt of it was … well, we were the black sheep of the family.'" </p> <p> As opposed to the Black
159. brmags/UK bottom of our ears."<LTH> We were the black sheep of the community," adds Melon bassist Brad.<LTH> Chr
160. brmags/UKen we've got it." 1977 <LTH> We're the black sheep of the New Wave." 1977 <LTH> I feel that the Mod sce
161. brmags/UKn to the City, its rebellious, wayward black sheep of a cousin - just West enough to keep the staid fro

162. npr/US My favorite track would be probably the Black Sheep one. <p> Excerpt of music from Black Sheep) <p> Hans
163. brbooks/UKanization that picks up an occasional black sheep, one only takes exception if it retains an accumulat
164. times/UK avoided him. `I know now that I'm the black sheep. Only three French riders have spoken to me since th
165. indy/UK ground." <p> And are you going to be a black sheep or a white sheep?" I ask Raphael, anxious to bestow
166. brmags/UK round a Swiss chalet-like cow shed. 3 Black sheep outside their Moorish palace home. 4 A Strawberry Hi
167. today/UK ton, now part of Scottish & Newcastle. Black Sheep raised &#pound; 750,000 in two weeks under a BES sche
168. usbooks/USover contempt with which the prodigal black sheep regards his erstwhile home town goes double for Some
169. usacad/US e on Elm Street. One of the rap group Black Sheep's favorite affectations is to mutter `Van Damme" (th
170. brmags/UK ighten up or let loose a little. Take Black Sheep's `State Of You". On a picnic, or to the park. Summe
171. econ/UK st temporary hiccups. `There are always black sheep," says Roger Fauroux, an inspecteur de finance who r
172. indy/UK nal families existed. Years later, when black-sheep sister LaToya spilled all the beans, and even Michae
173. brbooks/UKand without disloyalty. If Esau was a black sheep, so were all my best-loved friends - Ishmael and the
174. today/UK Family's example of how to convert the black sheep that lurks within most families. <p> Close friends
175. oznews/OZ favourite toy from Ireland was a tiny black sheep that cost Pounds Sterling1. She named him Kylemore
176. brmags/UKof others. We don't claim to be these `black sheep' that have f- all to do with anyone else.'" </p> <p>
177. times/UK ading cancer hospitals. <p> <h> Church black sheep;The Times Today </h> <p> The General Synod
178. oznews/OZtill deliberating on the case of their black sheep, the comedian Peter. Peter Berner, 36, is one of fou
179. guard/UK tance; among them `Dry River" and `The Black Sheep". The enamelled brilliance of the prose is remarkabl
180. brmags/UKmed horses.<LTH> Pieret was a talented black sheep - the son of a respectable Brussels attorney, who mi
181. brbooks/UKy," Donald Bakewell remarked. `Lot of black sheep there, I imagine. All safely in the past, of course.
182. brbooks/UK? Did there exist another world where black sheep thought of their erring lambs as the white sheep of
183. oznews/OZ VIVES EARLY TURBULENCE </hd> From the Black Sheep to the Judge Advocate General Corps -- ex-Marine Don
184. guard/UK THE transformation of Tottenham from black sheep to dark horses continues apace. Last night's narrow
185. usbooks/US to art in her family because of her `black sheep" uncle who was a depressed artist with a violent
186. usbooks/US as a sacred cow, by 1970 resembled a black sheep. Urban critics had argued early on that a rail track
187. usbooks/US family she might seem the rebellious black sheep, when Janet compared herself to friends like Nancy
188. times/UK Palumbo had been the invisible boy, a black sheep who had fled the white mansion Buckhurst, set in 60
189. oznews/OZ dings from their research are family `black sheep" who were murderers, or other criminals who served 1
190. guard/UK sort of handsome young man; charming black sheep who are irresistible to elderly women. She really
191. guard/UK fact that it's Michael, the disgraced black sheep, who also does the most to placate his father and
192. usbooks/USngarians and Chinamen" were ignorant `black sheep" whom industrialists could easily manipulate and use
193. brmags/UK ook by making much of it in an undyed black sheep wool.<LTH> I've always been fascinated by the gramma
194. usbooks/US herself like him, misunderstood, the black sheep. You can see their different ways of being in their
195. usephem/USnumber--> <!--and--> <!--price--> <h> BLACK SHEEP &#bull; A Wolf In Sheep's Clothing.</h> <p> Funky du

APPENDIX 2: non-canonical families

i) specification and addition

1. Two other brothers are rich and successful. <p> And that just leaves the runt of the litter, the black sheep of the Douglas family, saddled with the decidedly un-glamorous name of Eric and currently starring in a dead-end role of his own making - in a California prison cell.
2. EastEnders breaks it this week when a different Billy-Mitchell this time-appears in the Vic. The sneering, scowling drunk might be up against some stiff opposition but he still manages to be the most unpleasant black sheep of the Mitchell family.
3. Malcolm, "Doctor" (the Action Man) and Duncan, (the black sheep) of the "X family" all seem to have a thing about posing with guns.
4. 27 Charteris was the black sheep in Haig's family - the antithesis of the "gentlemen" on the staff.
5. Elvira-the camp vamp who looks like a black sheep from the Addam's Family.
6. At best, Spain is seen as the black sheep of the European family because, it is said, it did not have that obligatory rite of passage, a bourgeoisrevolution.
7. Like the Wimbledon side of the 80s, Tranmere are the poor relations in a city built on football passion. But since former Liverpool hero John Aldridge took charge, the black sheep of the Merseyside family, has begun to fight back.
8. Barnaby Gaitlin is a loser, just short of 30. He's the black sheep of a philanthropic Baltimore family. He has an almost pathological curiosity about other people's lives and a hopeless charm which attracts the kind of angelic woman who wants to save him from himself.
9. THE black sheep of one of Britain's wealthiest families, escaped jail yesterday despite admitting possessing cocaine and crack cocaine.
10. The main character, Piper (excellent Peter Gowen), is a sculptor, a self-hating black sheep of a Protestant ascendancy family and a repressed homosexual attracted to Conor McDermottroe's likeable Craig, a young blacksmith from Enniskillen.
11. When he was eventually unmasked he admitted his real name was Thornton -- and lied that he was the black-sheep of the famous chocolate family and had changed his name by deed poll.
12. Gasoline was banished from Melbourne as the black sheep of his blueblood family but new trainer Gerald Ryan hopes he he can become a Perth Cup candidate in today's race.
13. John Amery, the black sheep of a distinguished family whose father was the Secretary of State for India, was high on MI5's list of renegades" who helped the enemy in the Second World War.
14. Hugo Williams perfectly embodies most people's stereotype of a poet. Black sheep of a famous thespian family, ageless matinee looks, sensitive and wan, he suffers from bronchitis and happily confesses to a lifetime's avoidance of hard work.
15. I remember drinking the demon arak and talking late into the night with my late friend, Nalin: black sheep, as he told me, not without pride, of a famous family, a man who, among other achievements, pioneered the plays of Jean Genet in Colombo.
16. Malena (written by Senel Paz, the writer of Strawberry and Chocolate) features Ariadna Gil in the title role. Malena, the black sheep of a noble family, takes many

years to prove her own worth, having grown up in the shadow of her twin sister Reina, who has qualities everyone loves.

17. Gavaskar's snooty disdain for the bish-bash game is still a common enough view, especially in this country: limited-overs contests are routinely disparaged as bastard versions of the original, as if they were the loud-mouthed black sheep of some fine and ancient family.
18. There are 'black sheep' in the Compositae family, as all the artemisias, as all the artemisias, most of which are neurotoxic, belong to it.
19. Katsy Mecom collection of Art Nouveau & Tiffany Glass. 100 Tiffany glass miniatures and other pieces, all from the furnaces of the jewellery family's black sheep.
20. Everyone in showbusiness has had something done." <p> And every royal family has its black sheep - but it is unclear whether LaToya will be pleased at the news from Jermaine that there is room for her at Jackingham

ii) non-standard 'families'

1. FEDERICO ZERI was at once doyen, black sheep and chief scourge of the Italian art-history world, building up his huge international reputation through painstaking research
2. Roger Banks-Pye seems to be the black sheep amongst the Colefax and Fowler decorators.
3. His 200-page report called on the temple to stop its aggressive marketing campaign, and concluded: 'We do not want any black sheep in Thai Buddhism.'
4. There were no black sheep in Andersen - he would have found the idea distasteful. (You can't count the Ugly Duckling, for he was really a swan.)
5. The focus is on the sources of incomes earned by a herd of unusual and anomalous 'laborers', the black sheep of economic science: handmaidens and handymen, opera singers and opera dancers, priests and bureaucrats, kings and ministers, doctors and lawyers, leveraged buyout specialists, entrepreneurs and managers,
6. Maradona had been mooting the idea of a players' union for a long time but he said it had eventually materialised after talks with former French captain Eric Cantona, another black sheep of international soccer.
7. Indeed, Peter Rabbit in his own miniscule way is one of the true black sheep of literature. Like Alcott's Dan, he retains his integrity against all odds, refusing firmly to conform, withstanding every genteel effort to sickly him o'er with white.
8. THE life assurance industry - for long the black sheep of Britain's financial services - is struggling to mend its ways.
9. If French PM Lionel Jospin lifts the boycott, he risks a bitter backlash from his own farmers who want to keep our beef out. If he refuses to comply, France will become the black sheep of the EU.
10. Nynex, which many analysts consider the black sheep of the Baby Bells, might be in line for another shearing.
11. 'We were the only people there who grew hair past the bottom of our ears.' <LTH> We were the black sheep of the community," adds Melon bassist Brad.
12. The return of Ozawa: China's leader arrived in Japan just as the black sheep of the country's ruling party returned to form a new a new coalition TOKYO

13. <h> HEADLESS CHICKENS </h> <p> BLACK sheep of the label. </p> <p> Headless Chickens developed from being a dark, mean, industrial-type band on their first EP to their current position as both the only band on flying Nun most people would want to dance to
14. Five years ago it bought all the shares in the failed company. As a result the black sheep of the lending sector is now at the heart of the financial establishment.
15. <LTH> We're the black sheep of the New Wave." 1977 <LTH> I feel that the Mod scene was very close to the punk thing; wholly youth-like going out with green hair.
16. The Marquess of Bristol: The Marquess of Bristol, black sheep of the nobility, died on January 10th, aged 44
17. I liked the look and feel of the things, the starkness of their beauty. And I felt, and feel, curiously attracted by their shocking reputation - 900 years as the black sheep of the marksmanship world, the wicked-upstart strain of the weapon family, the unfair machine of war, the sniper's first friend, the reification of mankind's sneaky inhumanity towards its fellows.
18. The 29-year-old Mercatone rider said that since the drugs scandal at last year's Tour de France, cyclists had become the black sheep of the sporting world.
19. It hopes to snap up more of the thousands of water businesses serving New England. The move came as Yorkshire Water, the black sheep of the UK water industry during the 1995 drought when its former executives advised angry firms to quit Yorkshire, revealed plans to rename itself the Kelda Group, after an Old Norse word for `source of water`.
20. Bruno, who faces American Jesse Ferguson at Birmingham's NEC on March 16, is now ready to scrape the bottom of another barrel - and go for the WBO title, the black sheep of the world heavyweight championship belts, if he beats Ferguson. <p>
21. `It was fun and flamboyant and exciting," says Joey Ramone who, long before the Ramones, could be seen tottering down Queens Boulevard in New York in thigh-length boots and a pink jumpsuit. `I was definitely the black sheep of Queens! I got a lotta dirty looks. You were always walking around on the defensive: like, what the f # you lookin' at?

APPENDIX 3: *blue in the face*

1. indy/03 Mr Gorbachev can issue decrees until he is blue in the face, but they are ignored or sabotaged on th
2. indy/03 the television trade magazine) until you're blue in the face and have no idea how systems work inside
3. indy/03 ate that this a safe technology until one is blue in the face and people simply will not believe it.
4. oznews/16 things # we've discussed them until we're blue in the face," she said. <p> The kids know every darn
5. oznews/16 Cash not the issue - really AS WE go blue in the face waiting for a Super League signing to
6. oznews/16 and celebrated Waltzing Matilda until 're blue in the face, it's time to take a long, hard look one
7. oznews/16 st do it, and do it, and do it until we're blue in the face." <p> That is not a message designed to
8. oznews/16 and let the ARL scream and sue until it is blue in the face. <p> As it is now, Super League players
9. oznews/16 ea and pumpkin scones until they were also blue in the face. <p> Frustrated campaign team leaders
10. oznews/16 They argued until they were all shades of blue in the face and then capitulated. <p> Today, the
11. brmags/10 tion spans. But they can talk till they're blue in the face with the Brendas, because we just turn
12. brmags/10 ooting and inevitable excess until they're blue in the face. But now is the time for the modest
13. brmags/10 t, can I? <p> They can shout until they're blue in the face # she announces, momentarily every
14. brmags/10 o players in the dressing-room till you're blue in the face, it doesn't always mean that
15. brspok/07 three or four days of arguing till I was blue in the face that it was worth the money of going to
16. brspok/07 en to loads of people # about it until I'm blue in the face and really I'm just very confused and
17. brspok/07 e indeed. So you can admit it till you're blue in the face but you get nothing back so <F02> So
18. brspok/07 . Yeah. <M01> You can tell him till you're blue in the face that the rules have changed <F01> Yeah
19. brspok/07as irrelevant. You can campaign till you're blue in the face but <F01> Mm # <F02> this is the way it
20. guard/13 reaks down he can criticise it until he is blue in the face, but it will require the services of a
21. guard/13 and ourselves. You can shout until you are blue in the face, but psychiatric beds are still being
22. guard/13 d can go on setting up agencies until he is blue in the face or any other part of his anatomy.
23. npr/05 ly: the activists who've shouted till they're blue in the face, AIDS counselors and workers, whose job
24. npr/05 hat you had to pull and tug till you were <p> blue in the face to get them over your ankles. `They're
25. npr/05 We're going to use the word reform till we're blue in the face. And we may not have reform when you
26. npr/05 ions that he claims he answered until he was `blue in the face." So this I do think--and I think
27. brbooks/04 nt of trying on their part `til they were blue in the face was ever going to change that child.
28. brbooks/04 n. We can ask about Wee Jimmy until we're blue in the face. You have. And they don't even have to
29. brbooks/04 man what she wanted. He tried `til he was blue in the face and as long as she asked nicely. You
30. brbooks/04 od to be true. I've soundproofed till I'm blue in the face, but the noise of a piano you cannot
31. brbooks/04 tut-tutted until they were red, white and blue in the face, but Revie was gone. `We had to act fast
32. brbooks/04 e. I talked to Gooch about it until I was blue in the face but it was all falling on deaf ears. In
33. usbooks/12ntually, after Gretchen had talked herself blue in the face, pointing out all the marvelous
34. usbooks/12 talks about aluminum honeycomb until he's blue in the face. And it's <f> so <f> boring!" <p> It may
35. usbooks/12 er to Montpelier. Claire's talked herself blue in the face, but none of them'll give an inch."
36. usbooks/12 count and you can call them until you're blue in the face and they won't call you back. Some do.
37. usbooks/12 tov and the others talked until they were blue in the face. Ruskoi telephoned the military
38. usbooks/12 he results. Although we talk until we are blue in the face about the distinction between inquiring
39. usbooks/12 an tell yourself you're good until you're blue in the face, but until you get that kind of

40. usnews/11 anker said Friday. You can bid till you're blue in the face, but if management doesn't want you,
41. usnews/11 at the two sides could argue until they're blue in the face over whether firefighters should be
42. econ/08 ce dividend is peace, and argues until he is blue in the face that the Carter-Reagan defence build-up
43. econ/08 can (and frequently does) talk until he is blue in the face. Only a quarter of voters thought that
44. econ/08 concentrate on domestic policy until he is blue in the face; the truth of the matter is that in a
45. times/02 gs. I could lecture you like a mum till I'm blue in the face but the only people who can prevent
46. times/02 an call them Broadway operas" until you are blue in the face, but musicals they resolutely remain)
47. times/02 s", and added: `We have talked until we are blue in the face." <p> In Washington, Mr Clinton said
48. times/02 cedents from previous inquiries until he is blue in the face. When ministers have had eight days to
49. times/02 prepare all you want and putt until you are blue in the face, but if you get on the wrong side of
50. today/06 u can shout instructions at 'em till you're blue in the face, but they don't seem to listen.
51. today/06 trets. <p> I've read the Bail Act till I'm blue in the face and the magistrates misuse it. <p>
52. today/06 e Rangers-Leeds European Cup tie until he's blue in the face. <p> Unfortunately, it's just human
53. today/06 Teach </sh> <p> You can talk until you are blue in the face on the training ground but there is no
54. today/06 nd wear Gieves and Hawkes shirts till he is blue in the face, but nothing he does can make his
55. today/06 a physical training workload `until I was blue in the face. In the end I gave up because I
56. today/06 d people leaflets and offer advice till I'm blue in the face. Unless you, the sufferer, acts on the
57. today/06 ell the players what you want until you are blue in the face, but if they don't do anything about it
58. today/06 ELL would once have denied it until he was blue in the face. But nine albums into his career with
59. sunnow/17 he World: `The IRA can deny until they are blue in the face that they were involved, but we are
60. sunnow/17 e still not won. We've spoken until we are blue in the face about our tendency to fall asleep
61. times/02 . You can analyse football until you are purple in the face but, in the end, you are forced to
62. brmags/10 te strangers. Everyone's arguing themselves red in the face. Everyone's jabbing a finger at
63. guard/13 . <p> Kenneth Clarke can protest until he is red in the face - as he did earlier this week - that our
64. brbooks/04 his time. You could shout until you were black in the face then, and no bastard would hear you.
65. brbooks/04 ou may tell them not to talk till you're black in the face and they still do it # She - she looks
66. usbooks/12 . Three … # I counted until I was black in the face. At twenty I said, `Look, this isn't