Metaphor Annotation: A Systematic Study

(Technical Report CSRP-03-04)

A.M. Wallington, J.A. Barnden, P. Buchlovsky, L. Fellows and S.R. Glasbey
School of Computer Science, The University of Birmingham
Birmingham, B15 2TT, United Kingdom
A.M.Wallington@cs.bham.ac.uk

0.1 Introduction

This document describes a study of metaphor annotation that was carried out as part of the ATT-Meta project[1]. The study lead to a number of results about metaphor and how it is signalled that are reported elsewhere (e.g. Wallington et al 2003a, Wallington et al 2003b). It also resulted in a public database of metaphorical views (http://www.cs.bham.ac.uk/research/attmeta/DatabankDCA/index.html). The annotated files are also viewable on line at: http://www.cs.bham.ac.uk/~amw/dcaProject. The primary aim of this document is to describe how the project was set up and run, and to discuss the measures we took to identify and quantify inter-annotator (dis)agreement.

The study started in the summer of 2002, initially using seven paid data collection assistants (DCAs), but which stabilised at four after a couple of months, although one of the seven was able to rejoin the project in June 2003. During the period from July 2002 to October 2002 the DCAs were employed on a largely full-time basis, working on average between 30 and 40 hours a week. From October 2002 two of the DCAs had to restrict their hours to a maximum of 6 hours a week.

The texts used in the study came from two sources: a collection of doctor patient interviews about arthritis, and a selection of files from the British National Corpus. We shall refer to the work done on annotating the arthritis files as Phase 1 of the study and the work on annotating the BNC files as Phase 2. Altogether the files used in Phase 1 contained 65368 words and the files used in Phase 2 161689 words. This gave a combined corpus of almost a quarter of a million (227057) words. For training purposes prior to the start of the annotation exercise proper, a further set of arthritis files were used. These files were not included in the files for Phase 1. Furthermore, almost all the files were tagged separately by more than one DCA and for a number of files, two DCAs collaborated to arrive at a consensus annotation. This latter situation results in three versions of the file, one by each DCA plus a jointly tagged file. Having two or more annotations of the same file provides a useful insight into the degree to which trained annotators can differ in what they consider metaphorical. However, there are many advantages in allowing collaboration and discussion amongst annotators (see for example the work of the Pragglejaz group http://letlx.let.vu.nl/project/pragglejaz/) and we would have liked the majority of the files to be jointly tagged. Unfortunately, the nature of the hours worked by the DCAs, particularly from October 2002 onwards, restricted the amount of collaboration possible.

The DCAs working on the study were chosen as a result of a selection procedure and were restricted to candidates with English as their first language. The candidates were first given a short piece of text for training in which the metaphorical expression had been marked. They were then given an unannotated passage and asked to indicate the metaphorical stretches. During a following interview, the candidates were asked about metaphor and about the stretches of text they had or had not marked. There was a degree of flexibility with respect to the number of DCAs that could be employed and our aim in choosing DCAs was to select those who showed sensitivity to metaphor. Therefore we were confident in the abilities of the chosen DCAs to spot the majority of metaphorical stretches in a passage after training. This is not to say that there is always a right answer as to what words constitute a metaphorical stretch and weekly meetings and email discussions about some of the more uncertain stretches took place throughout the period of the study. As part of the study a means of testing inter-annotator agreement was devised. This will be described in section 4.

The mark-up of the text and the presentation of the data extracted from the text were designed to be XML compliant with a dtd. They were checked with the XML-Spy validator.

1 The ATT-Meta project was funded by the EPSRC, grant reference M64208/01, between 14/09/1999 and 13/06/2003

0.2 What was Annotated and How.

The files given to the DCAs to mark-up or annotate consisted of simple text files. In the case of the BNC files, the SGML mark-up was stripped out by running *sgml2txt* over the files. The process of annotation consisted of placing appropriate tags around the metaphorical stretches in the text and adding a list of the stretches that have been found, together with information of various kinds about the stretches to the end of the document.

0.3 identifying metaphorical stretches

In the initial set of instructions given to the DCAs, the DCAs were instructed to mark "interesting stretches" of text and were given instructions as to what counted as an "interesting stretch." The reason for using the term "interesting stretch" rather than "metaphorical stretch" is that one initial aim of the study was to investigate reliable means of identifying likely metaphors. Thus, for some of the DCAs the definition of an interesting stretch involved notions of "concreteness" and "abstractness" rather than metaphoricity per se.

The DCAs were divided into two teams: Team A and Team B. They were given the same set of files to annotate and asked to mark-up, what the study called, "interesting stretches". In the case of Team A, the definition of an "interesting stretch" made no reference to metaphor. The reasoning behind this is that the definition of what is a metaphor is a notoriously difficult to make. For this reason we hoped that the process of identification could be systematised into a number of steps involving the identification of simpler concepts. Accordingly, the following instructions were given to team A.

a word or phrase (a more precise definition will be given in section 1.2) is interesting if:

- its actual significance in the document is NON-PHYSICAL,
- in some other POSSIBLE context the sequence can have a PHYSICAL significance AND HAVE THE SAME SYNTACTIC NATURE,
- the physical significance is arguably RELATED to the abstract one.

In other words, we assumed that it would be easier to identify the concepts "physical" and "non-physical" than the concept metaphorical. This would help in the process of metaphor identification since many metaphors cast something abstract as something physical. It was recognised that certain types of metaphor go the other way or relate two abstract things, and so we accepted that these would be missed if the methodology of Team A were being followed.

So, in the sentence

"In the far reaches of her mind, Anne thought that Kyle was disloyal"

the noun phrase "far reaches" has abstract significance, because it qualifies Anne's mental state, but in another context it could still be a noun phrase and have a physical significance (as in "the far reaches of England"), where that physical significance is arguably what underlies the abstract significance.[2]

Team B were instructed to identify what they thought were metaphorical stretches including those where something abstract is viewed as something physical, something abstract is viewed as something abstract, something physical is viewed as something physical, or something physical is viewed as something abstract.

Examples that Team B should pick up, but which would not be identified by Team A include:

"image metaphors", (Lakoff and Johnson, 1980) where two things are likened because of their physical appearance; e.g. talking of a road as a snake (or as snaking), or of the light glinting on ocean waves as strings of a harp. This type of metaphor can be broadened to include "perceptual similarity metaphors" in general, where the similarity could be of sound, smell, or whatever, as in "the saw groaned".

Another special class (where both sides are at roughly the same level of abstractness) is what can be called "proper name metaphor", as in

"Lowry is the Picasso of this country."

"The Gulf War was another Vietnam."

2 Note that we are not claiming that "far reaches" only has an abstract reading in the context of mind. The following metaphor was used by an announcer introducing a Radio 4 program -Home Truths- on the BBC on the 24.2.2004: "and now a journey into the far reaches of family life".

"They're all Einsteins in that class."

Method A (the method for Team A) was abandoned when it was realised that too much polysemy in general was being annotated. It appears that word-senses follow Zipf's Law with most common words allowing many different word-senses, including both abstract and non-abstract senses. For example, the word bank can refer either to a physical building or the more abstract financial institution (as well as the homonym river bank). Indeed, it is by no means clear that the notion of 'word sense' is well-defined, and that for any word there is a fixed set of word senses (see Kilgarriff 1997*)*

The practice of using different methods of finding stretches was stopped after the first five files of Phase 1 (i.e., the preAmp). Consequently, for the second set of arthritis files and all the bnc files, the method used to identify interesting stretches was the method for Team B. The next section describes the instructions given to the DCAs about the syntactic form of the interesting stretches.

0.4 form of the annotations

An "interesting stretch" is a sequence of one or more words in the text with the following properties:

- if it is a single word, it is a noun (common or proper), a non-auxiliary verb, an adjective or an adverb. Note that this means that we are ignoring the issue of metaphorical extensions of prepositions (see Lakoff 1987, Tyler & Evans 2003).
- if it is more than a single word, it is a contiguous sequence of words that may not cross one of the following phrasal boundaries and finish phrase internally:

  noun phrase

  verb phrase

  adjectival and adverbial phrases

  prepositional phrase

  clause and sentence

The DCAs were instructed to keep the interesting stretches as short as possible whilst still consistent with being interesting and with including as many contentful specifiers or qualifiers as possible belonging to the same source domain. So, if an original sentence is

"In the far reaches of her mind, Anne thought that Kyle was disloyal"

"far" would be included in a sequence with "reaches" because it has a meaning that further specifies the "reaches" and although the sequence does not comprise a full phrase, it does not run over the noun phrase boundary. However, "of" is not included because "far reaches of" crosses the PP boundary. We don't include "the" because "the far reaches" doesn't have any more contentful specifiers than "far reaches" does. We don't say the stretch is "far reaches of her mind" because it could not itself have a physical meaning and is therefore not interesting.

0.5 XML compliancy

As we mentioned in the Introduction, our annotation scheme was intended to be XML compliant. Consequently, at the very start of each file, a tag of the following form is added:

<atmet_notated phase="**…**" team="**…**" annotator="**…**" annodate="**d/m/y**" lastmod="**d/m/y**" origfile="**…**">

The value of the phase is either **1** or **2**, with 1 referring to the phase of the study in which the arthritis texts were used and 2 referring to the phase in which the BNC texts were used. For various purposes the DCAs were divided into different, and not always the same, teams, with the particular team indicated by a letter. Initially, the teams differed in whether they used Method A or Method B, as described in section 1.2. The annotator identifier is the DCA's first and last name. The date is the date on which the DCA finished the annotation in the initially specified period for doing it. The lastmod value is the date and time of last modification, and the origfile value is the pathname of the file containing the original document.

At the very end of the file, the following closing tag is added:

</atmet_notated>

Turning now to the text itself and the stretches that the DCAs consider interesting. At the beginning of an interesting stretch of text the following "opening" tag is inserted:

<atmet N="...">

where the "..." is filled by a whole number. Note the double-quote marks round the number. At the end of that stretch of text a "closing" tag is inserted:

</atmet>

Thus, if the original sentence is

"In the far reaches of her mind, Anne thought that Kyle was disloyal"

then the tagged version could be:

"In the <atmet N="17">far reaches</atmet> of her mind, Anne thought that Kyle was disloyal"

Tags are numbered consecutively, starting with 1, as they are found, but there is no requirement that they appear in numerical order in the document (although there are some readability advantages if they do so). Consequently, "missed" or overlooked stretches can be added to previously-tagged text without adjusting the existing numbering.

0.6 atmetnotes

After the end of the document text a section is created that starts and ends with the tags

<atmetnotes>

</atmetnotes>

respectively.

In that section, for each metaphorical stretch that has been identified, e.g., N="17", there is a segment starting and ending with the tags

<atmetnote N="...">

</atmetnote>

where "..." is the note number. This segment contains the comments on the tagged stretch identified by the number and includes the following information, structured as indicated below.

0.7 stretch

A "stretch" component identifying the metaphorical/interesting stretch is included. Thus, in the case of the "far reaches" stretch in an example above, if it were stretch number 17, the corresponding atmetnote would be of form

<atmetnote N="17">
<stretch> far reaches </stretch>
</atmetnote>

0.8 function

A "function" element indicates what the interesting stretch is doing in the sentence, e.g. qualifying Anne's mental state, in the Anne/Kyle example. Thus:

<atmetnote N="17">
<stretch> far reaches </stretch>
<function> qualifying Anne's mental state </function>
</atmetnote>

0.9 confidence

An element giving your degree of confidence that the stretch is being used metaphorically.

<metaphorical> ... </metaphorical>

where the '...' is a whole number on a -2 to 2 scale with -2 meaning the annotator is certain that the stretch is not metaphorical, +2 meaning the annotator is certain that is metaphorical, and 0 meaning the annotator is undecided. Note that the number measures the level of confidence the DCA has that the stretch is metaphorical, not the degree of metaphoricity it has.

This might appear to be a somewhat odd attribute to ask for. Why annotate a stretch if it is not

metaphorical. The answer is, as we have said, that we originally tried two different methods of searching for metaphorical stretches, one of which makes no reference to metaphoricity. It was for this reason that the negative numbers on the scale were included. This method was abandoned after the first five files and for the rest of the study, the DCAs were asked to look for just metaphorical stretches. Consequently, the negative numbers were not required, although the option was kept. It was decided that the "metaphoricity" attribute was worth keeping since it made it more likely that the majority of metaphors would be spotted, albeit at a cost of including some possible non-metaphors amongst those stretches tagged with a 0 or 1. In total, 3 stretches were given a rating of '0' after a uniform definition of an interesting' stretch was given.

In the initial stage of the study, for those DCAs not looking directly for metaphorical stretches, a value of –2 or –1 would complete the section. The other DCAs and all the DCAs after the uniform definition would never need to use a negative number and so would also be required to add additional attributes.

0.10 metaphorical view

An attribute is included that gives the specific metaphorical view or conceptual metaphor that the stretch represents. If the tagged stretch is about a mental state, the metaphorical views may be listed in the ATT-Meta databank. The DCAs were also instructed to look in the Master Metaphor List (http://cogsci.berkeley.edu/). If the view could not be found in either of these two sources, then the DCAs were instructed to invent their own views. Periodically, the DCAs held meetings amongst themselves to draw up a list of views to be used in the study and this list was one of the principal outcomes of the project.

http://www.cs.bham.ac.uk/research/attmeta/DatabankDCA/index.html

This list includes views from the ATT-Meta databank, from the MML, and also those that the DCAs have invented. The list labels each view with its origin. With the existence of this list, the DCAs now had a fourth source of views, i.e. those created by the other DCAs. A fifth possibility is that the stretch is an idiom and the DCAs do not wish to analyse it as belonging to an existing view.

The DCAs were also asked to express their confidence that the stretch should be analysed as representing the stated view. As with all other confidence values in the project, a score of –2 should indicate that the annotator is confident that a particular property **does not** hold and a score of 2 should indicate that the annotator is confident that a particular property **does** hold. Thus, there should be no negative confidence values for this aspect of the annotation. However, it appears that some of the annotators misunderstood the instructions and a number of negative confidence values can be found. We should also point out that the means of recording the confidence level changed during the course of the project. Thus confidence can be included within the <view> label, but it was also marked within a separate <confidence> label.

<view from="**atmetdb**" confidence="1">COGNIZING AS SEEING</view>
<view from="**mml**" confidence="2">LIFE IS A JOURNEY</view>
<view from="**self**" confidence="**2**">POSSIBILITY IS ACCESSIBILITY</view>
<view from="**team**" confidence="**2**">COMMUNICATION IS A PATH</view>
<view from="**idiom**" confidence="2">IDIOM</view>

It is of course possible that a stretch should be viewed as combining two or more metaphorical views. For example, the phrase *the idea lay at the back of his mind* employs the view of MIND AS PHYSICAL SPACE and also IDEAS AS PHYSICAL OBJECTS. Consequently there was no requirement that only one view be given.

0.11 novelty

An important aspect of the project was to determine the extent to which metaphorical stretches can be seen as manifestations of conventional metaphorical views or are original. Thus, the DCAs were asked to add an element giving their degree of confidence that the tagged stretch rests on some completely novel metaphorical view (i.e., novel likening of two domains).

<novel> … </ novel>

where the ... is a whole number on a -2 to 2 scale with -2 meaning the DCA is very sure there is no novel view, +2 meaning the DCA is very sure there is, and 0 meaning the DCA is undecided.

0.12 novel variants

It is possible for a metaphorical stretch to rely upon a conventional metaphorical view, but to extend it

to bring in elements that are not a part of , i.e. to create not a novel metaphor, but a novel variant of a conventional metaphor. Consider for example, the following minimal pair.

She had an idea in mind.

She had an idea in the far reaches of her mind.

Both metaphors depend on the conventional metaphorical views MIND AS PHYSICAL SPACE and IDEAS AS PHYSICAL OBJECTS. But arguably, the second introduces an element that is not a part of the conventional view, namely 'the far reaches'. In the ATT-Meta project, elements, such as these, which are not considered to be part of a conventional mapping are called 'map-transcending elements'. The argument for considering this phrase a novel or map-transcending element rests largely on considerations of parsimony. Consider what the addition of the phrase 'the far reaches' adds to the information conveyed in the first sentence. We might say that in comparison to the degree of accessibility of the idea in the first sentence the second sentence expresses the view that the idea can only be considered or mentally manipulated to a limited degree. We find such inference patterns, where the conclusion reached is that something holds or is possible to a particular degree, across a wide range of metaphors, involving many different source and target domains. We would be missing a generalization if we were to tie this particular inference pattern to metaphors in which mind is viewed as physical space. Consequently, we would not want to include a specific mapping within the set of mappings comprising the MIND AS PHYSICAL SPACE metaphorical view, for distant parts of the mind as physical space, even if we could decide on a suitable analogue within the target concept of something as nebulous and poorly understood as the mind. Instead, we will argue that one of the ways in which metaphors can be extended as by introducing new elements that will allow specific inferences to be drawn about elements of the source domain for which there are conventional mappings. What these new elements are, of course, will depend on the particular dominant metaphorical view. One type of recurring inferences concerns the degree to which something holds. The issue is discussed further in Wallington et al 2003b.

After an early analysis of the results of the DCAs annotation, it was realised that there was no means of coding these 'novel variants' that allow conventional metaphors to be extended to bring in new inferences. Consequently, a new label of <novelvariant> was created with the '…' to be interpreted as usual.

<novelvariant> … </novelvariant>

Since this label was added part of the way through the project, not all the tagged files will contain it.

There are a number of problems with novel variant ratings:

1) When the metaphor is accorded a 'general' view like PERSONIFICATION or IMAGE, the novel variant is necessarily -2, since the view does not have any boundaries, and therefore nothing can be novel in the context of such a general view.

2) Since the novel variant field is rated within the context of the view, the statistics could be influenced if the view was not very good – i.e., if an imprecise view was used, the novel variant would be likely to be higher since the view would not accurately summarise the stretch.

There are also more general problems that hold of both the novel attribute and the novel variant, namely that there was no agreed set of metaphorical views, or within the metaphorical views any agreed set of mappings. Furthermore, in order to decide that a view or a source domain entity should be considered as novel, considerable careful argumentation and comparison with other views/mappings is required. Consequently, we found there was a tendency to minimize the degree of novelty.

0.13 claims

The DCAs were asked to indicate, using a number of labels, particular aspects of the information conveyed by the metaphorical stretch. This includes such information as to whether the metaphor is making a claim, the modality of the expression, whether it is part of someone's beliefs or emotions, or whether it conveys a value judgement. In this subsection we shall discuss the former, discussing the other labels in the following sections.

The DCAs were instructed to indicate whether the state of affairs that the stretch describes (or that the most immediately surrounding descriptive phrase describes) is being claimed to be the case, rather than being queried, commanded, stated to be a desire or hope, etc. The claim can be ascribed to someone else's mind, rather than being a claim of the author/speaker.

For example, the stretch "pulverizing/pulverize[s]" is not being used as a claim in the following:

Is Mary pulverizing the opposition?

John wonders whether Mary is pulverizing the opposition.

Mary, pulverize the opposition!

If Mary had pulverized the opposition John would have been pleased.

Mary wants to pulverize the opposition.

On the other hand, the stretch is being used as a claim in the following examples:

Mary is pulverizing the opposition.

Mary may be pulverizing the opposition.

John believes that Mary is pulverizing the opposition.

John says that Mary is pulverizing the opposition.

Mary says that she is pulverizing the opposition.

Because Mary is pulverizing the opposition, John is pleased.

When Mary pulverizes the opposition, John is pleased.

To indicate whether a claim is being made, the following label is used:

<claimlike> ... </claimlike>

with the '...' indicating the level of confidence as in the other cases.

0.14 within-mental

The DCAs were also instructed to add an element stating their degree of confidence that the stretch is within wording that displays the content of some mentioned person's thoughts or emotions. Such wording would be the parts in **bold** of the following sorts of things:

John believes [that] **Mary is a lemur**.

Their worry that **Mary is a lemur** ...

where the tagged stretch will be just "lemur"'. Thus, the labels would be as follows:

<within-mental> … </within-mental>

where the '...' is the level of confidence as usual. Note that for those metaphorical stretches that are not within-mental, the level of confidence would be a –2, i.e. an indication that the DCA is confident that the stretch is not within wording displaying the content of someone's thoughts or emotions.

0.15 modality

A further label is added to indicate if the stretch, or closely surrounding context, expresses negation, uncertainty or a modal qualification, using verbs, adverbs, etc. like

must, may, should, ought, not, not the case that, {suffix}n't, can, cannot, necessarily, possibly, maybe, might, [un]likely, probably

If it does, an element of the following form is included:

<qual> ... </qual>

where the '...' is the word[s] or phrase that conveys the qualification, e.g.,

<qual>wouldn't</qual>.

If modality is not expressed then the label is included, but given no value, i.e.

<qual></qual>.

0.16 value judgement

Another attribute of the stretch the DCAs were instructed to give is their degree of confidence as to whether the stretch conveys a judgement about the GOODNESS, IMPORTANCE or DESIRABILITY of what is being talked about is given. The judgement can be by the speaker/author, or by some agent mentioned in the text. For example, in the following sentence, an idea is cast as an object using the metaphorical view IDEAS AS PHYSICAL OBJECTS. Describing it as a gem conveys the speaker's

attitude towards the idea.

That's a gem of an idea.

As usual, this information is expressed in the following manner with the '…' indicating the degree of confidence as in other Subtasks (NOT the degree of goodness or whatever).

<vj> … </vj>

0.17 a complete atmetnote

With all the different labels completed, a finished atmetnote should be similar to the following:

<atmetnote N="**5**">
<stretch>over time</stretch>
<function>indicating time period</function>
<metaphorical>2</metaphorical>
<view from="**mml**" confidence="**2**">TIME IS A LANDSCAPE</view>
<novel>-2</novel>
<novelvariant> -1</novelvariant>
<claimlike>0</claimlike>
<within-mental>-2</within-mental>
<vj>-2</vj>
</atmetnote>

0.18 nbnotes

An additional system of notation is used for marking interesting things that the DCAs feel should not be included in the main tagging, but are, nonetheless, intriguing for some reason. As with the atmetnotes, an interesting stretch in a file is tagged and then information about the stretch is included at the end of the file. Thus the syntax of in-text tags is the following:

<nb N="..."> </nb>

and the format of the end-notes is:

<nbnote N="...">...any comments you like...</nbnote>

0.19 xref

When a stretch is very similar to one already tagged, whether in the same document or not, and all the elements in the note (other than the stretch element itself) would be exactly the same as that for the earlier stretch, it was decided that a cross-referenced atmetnoteref[3] should be used that contains only the note number and a cross-referencing element.

If cross-referencing within a document, the note should have the following form:

<atmetnoteref N="28" xref="27" />,

and, if cross-referencing to another document, the note should have the following form:

<atmetnoteref N="87" xref="39" doc="5.txt" />.

The in-text notation remains the same. Thus, <atmetnoteref N="28" xref="27" /> will refer to the stretch of text that has been marked up with, for example, <atmet N="28">**in the fight**</atmet>, but all the elements in the note will be identical to those of the atmetnote N="27" for <atmet N="27">**in World War Two**</atmet>.

0.20 prepositions and STANDARD tags

Owing to the frequency of certain prepositional phrases containing an abstract noun phrase (e.g. 'in trouble'), it was decided to introduce a new ATT-Meta "standard" tag to convey that the use of 'in' here is the standard use of the "container" view. Thus, for common prepositions, the atmetnotes would simply say "standard" rather than give information about the view - although other information such as "claimlike" would still need to be included.

---

3 Initially the same label 'atmetnote' was used for both proper atmetnotes and atmetnoterefs.

0.21 Later changes

It is likely that the DCAs, in the light of their experience with later files, may want to make changes both to what they consider to be an interesting stretch and to some of the elements within the atmetnotes. We considered it valuable to have a record both of the final considered version of the tagged text and of the prior stages before reaching the final version. Therefore, rather than allowing changes to overwrite the previous version, we decided to include in the annotation a record of what changes could be made as follows. Note that this applies only to "atmet" tags and atmetnotes. We decided that the "nb" tagging can be changed at will without making records.

1) Note Numbers

Once you have annotated a document, do not change any note numbers.

2) Deletions

If you want to undo the tagging of a stretch, delete the beginning and ending in-text tags, but keep the corresponding atmetnote, but change the start-tag of the end-note to the following:

<atmetnote N="..."
deleted="d"
startoffset="s">
</atmetnote>

where the “d” is the deletion date and time (to the nearest 10mins) together with a reason for the deletion, and “s” identifies the line in which the un-tagged stretch starts. This identification is the line number of the stretch start minus the line number of the first line of the tagged text (i.e. passing over the extra information at the top of the file

3) Additions

If you want to tag a previously untagged stretch, tag as normal except that you should change the start-tag of the end-note to the following:

<atmetnote N="..." added="a">

where the “a” is the addition date and time (to the nearest 10mins) together with a reason for it.

4) Minor Modifications to the Extent of a Tagged Stretch

If you want to make a minor change to what words are included in a tagged stretch, without significantly affecting the nature of what you've noticed in the text, just move the starting and/or ending in-text tags as appropriate, but add an attribute to the starting tag to get:

<atmet N="..." changed="c">

where the “c” is the date, time and reason for the change; if you need to change again, add a new "changed" attribute rather than changing or deleting the old one.

5) Modifications to the Information in an End-Note

If you want to make changes to the content of an end-note (other than to the stretch), then, within the relevant element, e.g. the novelty element, include a "changed" attribute to make the tag look, for instance, like the following:

<novel changed="c"> 2 </novel>

where the “c” gives the date, time and reason for the change.

If you need to change again, add a new "changed" attribute rather than changing or deleting the old one.

0.22 Metaphoricity Signals

Apart from annotating the texts for metaphorical stretches, another major part of the study was to investigate whether and how some metaphorical stretches are signalled in the text. Thus, we attempted to record the type and frequency of these 'metaphoricity signals', and how frequently they were used to signal a metaphorical stretch. Some early results of this study are reported in Wallington et al 2003a. In this paper we shall simply report how the DCAs identified and marked the metaphoricity signals.

The DCAs were given a document containing a list of metaphoricity signals based on work by Goatly

(1997), http://www.cs.bham.ac.uk/~amw/dcaProject/metaphoricity-signals.html. This document groups the signals into 12 classes, labelled A to L.

The DCAs were then told to search each document for all tokens of the metaphoricity signals listed and to tag each metaphoricity signals that was found with a new kind of tag, called a metsig.

Then, for each metaphoricity signal found, the DCAs were instructed to do the following:

1. At the beginning of a metaphorical signal stretch, put the opening tag <metsig N="?"> where the ... should be filled in by a whole number. number, as for the atmet.
2. At the end of a metaphorical signal stretch, put the closing tag </metsig>
3. The instructions for numbering metsig tags are exactly the same as those for atmet tags.
4. Create a section AFTER THE END of the atmetnotes section of the document, that starts and ends with the tags <metsignotes> and </metsignotes> respectively.
5. In the metsignotes section, for each note number as above, include a segment starting and ending with the tags <metsignote N="?"> and </metsignote> where ? is the note number. This segment contains your comments on the tagged stretch identified by the number, and should be structured as shown below.

    <metsignote N="17">
    <stretch> a bit of a </stretch>
    <sigclass> [A-L] </sigclass>
    <sigmet> -2 to 2 </sigmet>
    <signals> ? </signals>
    </metsignote>

The sigclass element should indicate the class of signals the stretch belongs to, where 'A' corresponds to the first of the classes in the list of signals, i.e. *Explicit signals of metaphoricity, similarity, etc. (where the signals themselves are not (very) metaphorical)*, and 'L' corresponds to the final class, i.e.*Commonization of Proper Names*.

The sigmet class indicates your degree of confidence that a metaphorical stretch is being signalled.

The signals element should indicate which (if any) stretches are signalled by the metaphorical signal in question. E.g. if the metaphorical signal “a bit of a” signals the stretch “goldmine”, which has been placed within an atmet tag, as in:

    <atmet N="91"> goldmine </atmet>

then the signals element of the metsignote should look as follows:

    <signals> 91 </signals>

Note that many tokens of metaphorical signals will not signal any stretches because the phrase in question is not actually being used as a metaphorical signal in this case. In these cases, the signals element should be left blank. These will always correspond to the cases in which the sigmet class is given a '-2' rating, and often when the sigmet class is '-1'.

Some tokens of metaphorical signals may signal more than one interesting stretch. In this case, include more than one signals element:

    <signals> 91 </signals>
    <signals> 92 </signals>

where "91" and "92" are the note numbers of the corresponding atmet tags.

0.23 Tools for Tagging

The task of tagging the texts was considerably helped by the use of various tools to insert tags and so on that were developed during the project. These mainly consisted of various emacs macros that could be loaded into the emacs text editor, i.e., the editor that was used for reading and editing the files.

In particular a main tagger (main-tagger.el) was developed with nine distinct functions for assisting in the tagging of texts. For example, two functions deal with tagging a metaphorical stretch. Firstly, the section can be highlighted, and the function tag-region used, or a single word can be tagged

automatically using the function tag-word.

1. Tag-region

This function presumes an input of a region, it then calls begin-tag which inserts

<atmet N="(the counter)">,

moves to the end of the region and calls, end-tag, which inserts

</atmet>

moves to the end of the file and calls add-base, which inserts

<atmetnote N=" (the counter) ">
<stretch> (the piece of tagged text)</stretch>
<function></function>
<metaphorical></metaphorical>
<view from=""></view>
<novel></novel>
<claimlike></claimlike>
<within-mental></within-mental>
<qual></qual>
<vj></vj>
</atmetnote>

2. Tag-Word

This function simply calls tag-region, although setting the tagged region to a single word. This presumes no highlighted region input, but requires the cursor in emacs to be positioned immediately before the desired word to tag.

3. Function Three - Tagging a note

This function works in the same way as Tag-Region, but using

<nb N = "(counter)">, </nb>

and

<nbnote N="(counter)"> </nbnote>

to mark-up the in-text and post-text parts of the file respectively.

4. Function Four - Tagging a reference

This uses an identical procedure as tag-region, although inserting

<atmetnote N=" (counter) "xref= "…"> </atmetnote>

as the base tag.

5. Function Five – STANDARD tags

This too uses an identical procedure as tag-region, but inserts

<analysis>STANDARD</analysis>

as the base tag.

6. Function Five - Tagging a metaphoricity signal

Again, this works in an identical manner to tag-region, but using the label 'metsig' and

<metsignote N=" (counter) ">
<stretch>" stretch "</stretch>
<sigclass></sigclass>
<sigmet></sigmet>
<signals></signals>
</metsignote>

7. Function Six - Setting the counter

The counter is initially set at 0, and incremented every time one of the tag procedures is called, before it outputs. Note that it needs to be reset every time emacs is opened when continuing a new file, or when moving on to metaphoricity signals, as they currently use the same counter.

8. Function Seven - Adding the finishing tags

The final stage of tagging a document requires information to be inserted at the beginning and end of a document. From any location in the file, this inserts

```
<atmet_notated
phase=""
team=""
annotator=""
annodate=""
lastmod=""
origfile="">
```

at the beginning and

```
<atmetnotes>
```

at the end.

9. Function Eight - Highlighting the tags

This simply places emacs into sgml mode, which colors the tags.

These functions were mapped as follows:

F3 tag-signal
F4 tag-reference
F5 tag-word
F6 tag-region
F7 set-counter
F8 finish-off
F9 add-note
F10 place in SGML mode
F11 tag STANDARD

0.24 Inter-Annotator Agreement.

As we stated in the introduction, the majority of the files were tagged by more than one DCA. This allows not only for a check to be made as to whether any of the DCAs were failing to spot many metaphorical stretches, but more importantly allows a study to be made of how trained annotators might legitimately differ in their decision as to what is metaphorical and what is not. Texts which have been tagged by a group of taggers attempting to reach a consensus (e.g. the Pragglejaz group http://letlx.let.vu.nl/project/pragglejaz/) give no indication of how much disagreement was encountered. Furthermore, one of the original aims of the project was to compare the results of using two different methods of tagging "interesting stretches".

A tool was created[4] that allows the extraction of the different attributes to determine the proportion or count of the metaphorical stretches and metaphoricity signals that possess a particular property, e.g. making a value judgement, or having a negative value (i.e. -1 -2). This information was designed to be viewed in a spread sheet. In addition standard unix scripting tools such as egrep, sort, and uniq were used to extract information from the tagged files. The gathering of such information from each of the DCAs allows comparisons to be made. However, using such methods only allows a comparison of the counts to be made. It is possible, although highly unlikely, that two DCAs could report identical numbers of metaphors, but disagree completely on what stretches should be tagged as being metaphorical. The situation is further complicated by the possibility that DCAs may agree on the core of a metaphorical stretch, but disagree on the start and/or finish of the stretch. We would consider there to be substantial agreement if the stretches one DCA tagged as metaphorical overlapped with all the stretches another DCA tagged as metaphorical. However, calculating the degree of overlap between the different DCAs is an intensive affair.

4 Written in java.

For this reason the following alternative method was chosen.

Note the following definitions:

- "word token" = occurrence of a word in the document (a given word such as "impact" can appear as many different tokens in the document)
- "tagged word token" = a word token in a tagged stretch.
- "DCA tagged word token" = DCA included the token in a tagged stretch

Then,

1) For each tagged word token successively in the whole document, such that the word token is tagged by more than one DCA, report:
   - the ordinal position of the word token in the document (e.g. 234th token)
   - the word itself
   - the number of DCAs who included it in a tagged stretch.

2) If the number of DCAs who took part in the exercise is N, report:
   - the number of word tokens that were tagged by N DCAs
   - the number of word tokens that were tagged by N-1 DCAs
   - .....
   - the number of word tokens that were tagged by 1 DCA

3) For each DCA, report:
   - the number of word tokens he/she tagged that are in the N group in (3)
   - the number of word tokens he/she tagged that are in the N-1 group in (3)
   - .....
   - the number of word tokens he/she tagged that are in the 1 group in (3)

This measure, as it is, will be inflated by an amount corresponding to the effect of the just randomly tagging words. Consequently we need to "kappa-adjust" the counts from (3).

4) To do this, we proceed as follows:
   - calculate P, the number of tagged word tokens in the document divided by the number of word tokens in the document
   - calculate S, the number of tagged stretches in the DCA's tagged document
   - for each DCA D.
   - For each K from 2 to N,
   - subtract the following from the count derived in (3) for the K group:
     - (the P value for DCA D) * (S)**(K-1)), (where ** means exponentiation.)
     - (the number of words in the conversation in the document)

Finally , the result of dividing the counts derived in (3) and the adjusted counts derived in (4) by the corresponding group sizes from (2) should also be reported.

References

Goatly, A. 1997. *The language of metaphors*. London and New York: Routledge.

Kilgarriff, Adam 1997 *Computers and the Humanities 31 (2), pp 91--113*

Lakoff, George. 1987. *Women, Fire and Dangerous Things*. University of Chicago Press.

Lakoff, G. & Johnson, M. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.

Tyler, Andrea and Vyvyan Evans 2003. *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition*. CUP.

Wallington, A.M., Barnden, J.A., Barnden, M.A., Ferguson, F.J. & Glasbey, S.R. 2003a. *Metaphoricity Signals: A Corpus-Based Investigation*.Technical Report CSRP-03-5, School of Computer Science, The University of Birmingham, U.K.

Wallington, A.M., Barnden, J.A., & Glasbey, S.R. 2003b. *View Neutral Mapping Adjuncts in Real Text: A Preliminary Investigation*. Technical Report CSRP-03-6, School of Computer Science, The University of Birmingham, U.K.