CMT-based Metaphor Annotation

Ekaterina Shutova

1 Introduction

The study of metaphor dates back to the times of Aristotle and touches on various aspects of human reasoning and multiple disciplines. Since the first inquiries, the theory of metaphor has evolved significantly under the influence of linguistic and psychological findings [3, 59, 30, 28, 24, 14, 4, 16, 17], and the establishment of the fields of artificial intelligence [2, 44], cognitive science [22] and neuroscience [12]. Following Aristotle's *Poetics*, it is widely acknowledged across these disciplines that metaphor is based on *analogy* [13, 30, 20, 44, 10] and arises when one concept is viewed in terms of the properties of another. Humans often use metaphor to describe abstract concepts through reference to more concrete or physical experiences. Below are some examples of metaphor.

(1) How can I *kill* a process? [37]

(2) Hillary *brushed aside* the accusations.

(3) I *invested* myself fully in this research.

(4) And then my heart with pleasure *fills*,
And *dances* with the daffodils.
("I wandered lonely as a cloud", William Wordsworth, 1804)

Metaphorical expressions may take a great variety of forms, ranging from conventional metaphors, which we produce and comprehend every day, such as those found in (1), (2) and (3), to poetic and novel ones, such as (4). In metaphorical expressions, seemingly unrelated features of one concept are attributed to another concept. In example (1), a *computational process* is viewed as a *living being* and, therefore, its

Ekaterina Shutova
International Computer Science Institute, University of California, Berkeley
e-mail: katia@berkeley.edu

forced termination is associated with the act of killing. In (2) Hillary is not literally clearing away the accusations with a brush. Instead, the accusations lose their validity in that situation, in other words Hillary *rejects* them. The verbs *brush aside* and *reject* both entail the resulting disappearance of their object, which is the shared salient property that makes it possible for this analogy to be lexically expressed as a metaphor.

Metaphor has traditionally been viewed as an artistic device that lends vividness and distinction to its author's style. This view was challenged by Lakoff and Johnson [30], who claimed that it is a productive phenomenon that operates at the level of mental processes (see also [48]). According to Lakoff and Johnson, metaphor is not merely a property of language, i.e. a linguistic phenomenon, but rather a property of thought, i.e. a cognitive phenomenon. This view was subsequently acquired and extended by a multitude of approaches [20, 44, 10, 12, 46] and the term *conceptual metaphor* was coined to describe it. Conceptual metaphor is not limited to similarity-based meaning extensions of individual words, but rather involves reconceptualisation of a whole area of experience in terms of another. Thus metaphor always involves two concepts or conceptual domains: the *target* (also called *topic* or *tenor* in linguistics literature) and the *source* (also called *vehicle*). Consider the following examples.

(5) He *shot down* all of my arguments. [30]

(6) He *attacked* every weak point in my argument. [30]

(7) Your claims are *indefensible*. [30]

(8) I *demolished* his argument. [30]

(9) I've never *won* an argument with him. [30]

(10) You disagree? Okay, *shoot*! [30]

According to Lakoff and Johnson, a mapping of the concept of *argument* to that of *war* is employed in all of these examples. The *argument*, which is the target concept, is viewed in terms of a *battle* (or a *war*), the source concept. The existence of such a link allows us to talk about *arguments* using *war* terminology, thus giving rise to a number of metaphors. Conceptual metaphor, or *source-target domain mapping*, is thus a generalisation over a set of individual metaphorical expressions that covers multiple cases in which one domain can be described using the language of another. However, critically, Conceptual Metaphor Theory (CMT) does not merely claim that cognitive metaphor provides a means for producing or understanding metaphorical expressions; rather, cognitive metaphor is viewed as an important organizing principle of the conceptual system. The systematic mappings of conceptual metaphors provide a cognitive mechanism for representing and reasoning about a target domain in terms of a source domain. The CMT therefore makes a very strong claim about how conceptual knowledge is organized in the mind. However, a key assumption is that there exists a fixed set of correspondences between pairs of domains that form the basis of the activated mapping.

Lakoff and colleagues put forward many examples of conceptual metaphor systematically supported by a set of metaphorical expressions found in language. According to them, manifestations of conceptual metaphor are ubiquitous in language and communication. Below are a few other examples of common metaphorical mappings, that are widely agreed upon.

- TIME IS MONEY (e.g. "That flat tire *cost* me an hour")
- IDEAS ARE PHYSICAL OBJECTS (e.g. "I can not *grasp* his way of thinking")
- LINGUISTIC EXPRESSIONS ARE CONTAINERS (e.g. "I would not be able to *put* all my feelings *into* words")
- EMOTIONS ARE VEHICLES (e.g. "[...] she was *transported* with pleasure")
- FEELINGS ARE LIQUIDS (e.g. "[...] all of this *stirred* an unfathomable excitement in her")
- LIFE IS A JOURNEY (e.g. "He *arrived* at the end of his life with very little emotional *baggage*")

Lakoff and colleagues further demonstrated their ideas in a resource called Master Metaphor List (MML) [29]. The list is a collection of source–target domain mappings (mainly those related to mind, feelings and emotions) with corresponding examples of language use. The mappings in the list are organised in an ontology, e.g. the metaphor PURPOSES ARE DESTINATIONS is a special case of a more general metaphor STATES ARE LOCATIONS. To date MML is the most comprehensive metaphor resource in the linguistic literature.

CMT produced a significant resonance in the fields of philosophy, linguistics, cognitive science and artificial intelligence, including natural language processing (NLP). It inspired novel research [38, 39, 44, 45, 2, 11, 41, 40, 1], but was also criticised for the lack of consistency and empirical verification [43, 49, 46, 42]. The theory relies on a very strong assumption that there exists a fixed set of cross-domain correspondences, and that it is possible to describe them using a fixed set of predefined domain labels. However, it is yet to be verified whether this assumption holds. The evidence commonly presented in support of CMT is based on introspections about a set of carefully selected examples, such as those in the Master Metaphor List. Such introspections, albeit clearly illustrating the main tenets of the theory, are not a satisfactory substitute for empirical data [42]. These examples cannot possibly capture the whole spectrum of metaphorical expressions in unrestricted, naturally-occurring text, and thus do not provide evidence that the theory can adequately explain all (or at least the majority) of metaphors used and dynamically created in real-world communication. An annotation study of conceptual metaphor in continuous text is needed for the latter purpose. This chapter presents a study, in which we use linguistic annotation techniques to verify whether all linguistic metaphors can be explained by a corresponding conceptual metaphor, and whether the labels can be consistently assigned to source and target domains.

Previous corpus-linguistic studies of CMT looked at metaphorical mappings within a limited domain, e.g. WAR, BUSINESS, FOOD or PLANT metaphors [23, 55, 8, 19, 36, 35, 21], in a particular genre or type of discourse [36, 23, 55, 21], or though individual examples in isolation from wider context [58, 33], often focusing

on a small predefined set of source and target domains. Despite the popularity and impact of CMT, there still has not been a corpus-based study covering all metaphorical expressions and their respective mappings in on-going, open-domain discourse, nor a comprehensive procedure for such annotation in free text. However, a general-domain corpus annotated for metaphorical associations could provide a new starting point for linguistic, cognitive and computational experiments on metaphor. The annotation scheme we present in this chapter is also a step towards filling this gap. It is a revision and extension of the pilot study we first presented in [51, 50]. We designed the annotation study to reveal (1) how intuitive the conceptual metaphor explanation of linguistic metaphors is for human annotators and whether it is possible to consistently annotate interconceptual mappings; (2) what are the main difficulties that the annotators experience during the annotation process; (3) whether one conceptual metaphor is sufficient to explain a linguistic metaphor or whether a chain of conceptual metaphors is needed; and (4) what proportion of metaphorical expressions can be explained using the proposed lists of most general source and target categories suggested in the MML.

The annotation scheme we developed is a joint scheme for identification of metaphorical expressions and source-target domain mappings. It thus addresses two problems: the distinction between literal and metaphorical language in text and the formalisation of human conceptualisation of metaphorical mappings. Rather than assume the existence of a set of pre-defined and fixed metaphorical mappings as claimed in CMT, the annotation procedure we adopted does not rely on such mappings, but instead makes use of independent sets of common source and target domain categories. Such a setting allows to test the CMT against the annotated corpus data. For example, if a given source domain is systematically mapped to the same target domain in the corpus (i.e. LIFE is always paired with JOURNEY) then this is evidence for a deep representational correspondence between these two domains and thus support for the CMT. However, on the other hand, if a given target domain tends to vary in how it is paired with source domains (both across linguistic inputs and across annotators) then this is evidence against the hypothesis that any particular source domain structures understanding of the target domain. The data thus provide a test of whether there is a restricted set of mappings which structure conceptual domains (as specified by the master metaphor list) or rather whether humans use an unrestricted range of different mappings and pair domains on a more ad-hoc basis that depends on the particular linguistic input.

The annotation was carried out on real-world texts taken from the British National Corpus (BNC) [5], representing various genres. We tested the scheme in an experimental setting involving multiple annotators and measured their agreement on the task. The focus of the study is on single-word metaphors expressed by a verb. Restricting the scope to verbs was a methodological step aimed at testing the main principles of the proposed approach in a well-defined setting and it was done without loss of generality. The choice of verbs was primarily motivated by their high frequency in metaphorical constructions, according to corpus studies. For example, [6] conducted a corpus study of the use of metaphor in educational discourse for all parts of speech. She found that verbs account for around 50% of the data,

the rest shared by nouns, adjectives, adverbs, copula constructions and multi-word metaphors. This suggests that verb metaphors provide a reliable testbed for our experiments. The annotators were asked to (1) classify the verbs in the text into two categories: metaphorical or literal and (2) identify the interconceptual mapping for each verb they tagged as metaphorical. For the second task, the annotators were given precompiled lists of suggested source and target domain labels, from which they selected the categories that – in their judgement – described the source and target concepts best. However, they were also allowed to introduce their own category if the relevant list did not contain the desired one. We expect the assignment of domain labels to be the most challenging part of the annotation process. The main goal of the study is thus to verify whether such labels can be assigned consistently, which could in turn provide evidence in support of CMT.

Only a part of the corpus was annotated by multiple independent annotators, to measure reliability. The rest of the dataset was annotated by one annotator only. Additionally, linguistic metaphors expressed by nouns, adjectives and adverbs were also annotated (in a single annotator study), in order to estimate metaphor statistics across part-of-speech classes and syntactic constructions.

The chapter first describes previous work on metaphor annotation, then our own dataset and annotation scheme used to identify both linguistic and conceptual metaphor in text, and finally concludes with the annotation reliability study conducted in a setting with multiple annotators and the analysis of the resulting corpus.

2 Previous approaches to metaphor annotation

The task of metaphor annotation in corpora can be split into two stages, to reflect two distinct aspects of the phenomenon, i.e. the presence of both linguistic and conceptual metaphor. These stages include the identification of metaphorical senses in text, which requires distinguishing between literal and non-literal meanings, and the assignment of the underlying source-target domain mappings. Although humans are perfectly capable of producing and comprehending metaphorical expressions, the task of annotating metaphor in text is challenging. This might be due to the variation in its use and external form, as well as the conventionality of many metaphorical senses. Gibbs [15] suggests that literal and figurative meanings are situated at the ends of a single continuum, along which metaphoricity and idiomaticity are spread. This makes demarcation of metaphorical and literal language fuzzy.

Traditional approaches to metaphor annotation include the manual search for lexical items used metaphorically [47], for source and target domain vocabulary [9, 25, 40] or for linguistic markers of metaphor [18]. Gerard Steen and the Praggle-jaz group [47] proposed a metaphor identification procedure (MIP) for human annotators. The procedure involves metaphor annotation at the word level as opposed to identifying metaphorical relations (between words) or source–target domain mappings (between concepts or domains). In order to discriminate between words used metaphorically and literally, the annotators are asked to follow the guidelines pre-

1. Read the entire text-discourse to establish a general understanding of the meaning.
2. Determine the lexical units in the text-discourse.
3. • For each lexical unit in the text, establish its meaning in context, that is, how it applies to an entity, relation, or attribute in the situation evoked by the text (contextual meaning). Take into account what comes before and after the lexical unit.
   • For each lexical unit, determine if it has a more basic contemporary meaning in other contexts than the one in the given context. For our purposes, basic meanings tend to be
     – More concrete [what they evoke is easier to imagine, see, hear, feel, smell, and taste];
     – Related to bodily action;
     – More precise (as opposed to vague);
     – Historically older;

     Basic meanings are not necessarily the most frequent meanings of the lexical unit.
   • If the lexical unit has a more basic current contemporary meaning in other contexts than the given context, decide whether the contextual meaning contrasts with the basic meaning but can be understood in comparison with it.
4. If yes, mark the lexical unit as metaphorical.

Fig. 1 Metaphor identification procedure of Pragglejaz Group

sented in Figure 1. In the framework of this procedure, the sense of every word in the text is considered as a potential metaphor, and every word is then tagged as literal or metaphorical. Thus such annotation can be viewed as a form of word sense disambiguation with an emphasis on metaphoricity. MIP laid the basis for the creation of the VU Amsterdam Metaphor Corpus[1] [56] (see previous chapter). This corpus is a subset of BNC Baby[2] annotated for linguistic metaphor. Its size is 200,000 words and it comprises four genres: news text, academic text, fiction and conversations.

Martin [40] conducted a corpus study in order to confirm that metaphorical expressions occur in text in contexts containing lexical items from source and target domains. The difficulty associated with this approach is that it requires exhaustive lists of source and target domain vocabulary. The analysis was performed on the data from the Wall Street Journal (WSJ) corpus [7] and focused on four conceptual metaphors that occur with considerable regularity in the corpus. These included NUMERICAL VALUE AS LOCATION, COMMERCIAL ACTIVITY AS CONTAINER, COMMERCIAL ACTIVITY AS PATH FOLLOWING and COMMERCIAL ACTIVITY AS WAR. Martin manually compiled the lists of terms characteristic for source and target domains by examining sampled metaphors of these types and then extended them through the use of a thesaurus. He then searched the corpus for sentences containing vocabulary from these lists and checked whether they contain metaphors of the above types. The goal was to evaluate the predictive ability of contexts containing vocabulary from the source domain and the tar-

[1] `http://www.ota.ox.ac.uk/headers/2541.xml`

[2] BNC Baby is a four-million-word subset of the British National Corpus (BNC) [5], comprising four different genres: academic, fiction, newspaper and conversation. For more information see `http://www.natcorp.ox.ac.uk/corpus/babyinfo.html`

get domain. In addition, Martin estimated the likelihood of a metaphorical expression following another metaphorical expression described by the same mapping. The most positive results were obtained for metaphors of the type NUMERICAL VALUE AS LOCATION ($P(Metaphor|Source) = 0.069$, $P(Metaphor|Target) = 0.677$, $P(Metaphor|Metaphor) = 0.703$). The low predictive ability of the source domain vocabulary may be due to the fact that source domains normally refer to our physical experiences. Consequently, the associated vocabulary would tend to occur independently and literally, as opposed to more abstract (target) concepts that frequently appear in metaphorical constructions.

[57] experimented with metaphor annotation in unrestricted text. They employed two teams of annotators and compared externally prescribed definitions of metaphor with intuitive internal ones. Team A was asked to annotate "interesting stretches", whereby a phrase was considered interesting if (1) its significance in the document was non-physical, (2) it could have a physical significance in another context with a similar syntactic frame, (3) this physical significance was related to the abstract one. Team B had to annotate phrases according to their own intuitive definition of metaphor. Apart from metaphorical expressions, the respective source-target domain mappings were also to be annotated. For this latter task, the annotators were given a set of mappings from the Master Metaphor List and were asked to assign the most suitable ones. However, the authors do not report the level of interannotator agreement, i.e. the proportion of instances that were tagged similarly by all annotators, nor the coverage of the mappings in the Master Metaphor List on their data. The fact that the method of [57] is limited to a set of mappings exemplified in the Master Metaphor List suggests that it might not scale well to real-world data, since the predefined inventory of mappings is unlikely to be sufficient to cover the majority of metaphorical expressions in arbitrary text.

3 Data

Our annotation study was conducted on a set of texts taken from the British National Corpus. BNC is a 100 million word corpus containing samples of written (90%) and spoken (10%) British English from the second half of the 20th century. The data for it was gathered from a wide range of sources and the corpus is balanced with respect to genre, style and topic. As such, it provides a suitable platform for the development of a metaphor corpus, aimed at the study of metaphor in real-world texts in contemporary English.

To collect the data for the metaphor corpus we sampled texts from the BNC representing various genres, aiming to retain the genre balance of the BNC itself to the extent possible. The data included fiction (5,293 words), newspapers (2,086 words) and journal articles (1,485 words), essays on politics, international relations and sociology (2,950 words), and radio broadcasts (transcribed speech, 1,828 words). This allowed for a study of metaphor in diverse discourse. The total size of the corpus annotated is 13,642 words.

4 Annotation scheme

Our task is to identify both linguistic metaphors and the corresponding conceptual metaphors. The annotation process will, therefore, operate in two stages. First, lexical items are classified as either metaphorically or literally used. Then, for all cases of metaphorical use the appropriate source-target domain mappings need to be assigned. The annotation scheme thus addresses two problems: the distinction between literal and metaphorical language in text and the formalisation of human conceptualisation of metaphorical mappings.

4.1 Main principles and challenges

The key desiderata in developing such a metaphor annotation procedure concern the choice of the level of conventionality of the metaphorical expressions to be annotated and a suitable inventory of source and target domain categories used to assign the mappings.

- **Level of conventionality** As already mentioned in section 2, the distinction between metaphorical and literal meanings is not always clear-cut. A large number of metaphorical expressions are conventionalised to the extent that they are perceived as literal by most native speakers (e.g. "He *found out* the truth"). Some approaches to metaphor consider only novel expressions to be truly metaphorical [27], whereas others consider any linguistic expression to be metaphorical where an underlying analogy can be identified [56]. In this study we consider both novel and conventional metaphors as interesting for annotation; however, we only include the conventional cases where both literal and metaphorical senses are commonly used and stand in clear opposition in contemporary language. This is where the scope of our annotation differs from that of [56], who is additionally interested in the historical aspects of metaphor.
- **Inventory of categories** The primary question one faces when trying to derive an annotation scheme for metaphorical associations is defining a set of source and target domain categories. As opposed to the previous approach of [57], who used a predefined set of fixed mappings from the MML (e.g. LIFE IS A JOURNEY), in our scheme both source (e.g. JOURNEY) and target (e.g. LIFE) domains can be chosen independently. We expect that this will allow for higher flexibility of annotation and thus provide a better reflection of human intuitive conceptualisation of metaphor, as well as the identification of novel mappings.
  The main properties of categories to consider while designing and evaluating such an annotation scheme are their coverage and specificity. The inventory of categories should cover a wide range of topics and genres. The categories themselves should be at the right level of generality, i.e. not too general (to ensure they are sufficiently informative for the task), but at the same time not too specific (to ensure they provide high coverage of the data).

The remainder of this section describes how the annotation scheme was developed and tested with these principles in mind.

4.2 Source and target domain categories

To date the most comprehensive resource of metaphorical mappings is the Master Metaphor List [29]. Its source and target domain categories were repeatedly adopted for linguistics and NLP research [2, 32]. Following these approaches, we relied on a subset of categories from the Master Metaphor List to construct the inventory of categories for annotation.

We selected a number of general categories from the MML, e.g. LOCATION, CONTAINER, JOURNEY, LIFE, TIME, RELATIONSHIP, and arranged them into source and target concept lists. These lists were then given as suggested categories to annotators. Suggested source and target concepts are shown in Tables 1 and 2 respectively. The expectation is that the categories in these lists would account for a considerable proportion of metaphorical data, i.e. provide a reasonable, albeit not exhaustive, coverage. In order to test their coverage, we conducted a pilot study on a small text sample (2,750 words) from the BNC. We annotated metaphorical expressions and the corresponding interconceptual mappings in these texts using the categories from the suggested source and target concept lists. The study revealed that the target concept list accounted for 76% of metaphorical expressions in these texts, whereas the source concept list had a 100% coverage. Such discrepancy can be explained by the fact that target categories, which tend to describe abstract concepts, are significantly less restricted than source categories that stand for our physical experiences. In other words, we can use metaphor to talk about an unlimited number of abstract things, whereas the entities, events and processes to which we compare them are limited to the actual physical experience we all share. Thus the set of potential target concepts is likely to be significantly larger and harder to predict. To account for this, the annotators, although strongly encouraged to use categories from the provided lists, were allowed to introduce novel categories in cases where they felt no category from the lists could adequately explain the instance. Since metaphor production and comprehension is open-ended by definition (i.e. novel metaphorical mappings can be produced and understood by humans), this step is crucial for annotation or real-world data. However, the suggested categories provide the necessary guidance on what such new categories may be like.

4.3 Annotation procedure

Metaphor annotation is carried out at the word level. The proposed annotation scheme is based on some of the principles of the metaphor identification procedure developed by [47]. We adopt their definition of a basic sense of a word and

Source concepts
PHYSICAL OBJECT
LIVING BEING
ADVERSARY/ENEMY
LOCATION
DISTANCE
CONTAINER
PATH
PHYSICAL OBSTACLE (e.g. barrier)
DIRECTIONALITY: e.g. UP/DOWN
BASIS/PLATFORM
DEPTH
GROWTH/RISE
SIZE
MOTION
JOURNEY
VEHICLE
MACHINE/MECHANISM
STORY
LIQUID
POSSESSIONS
INFECTION
VISION

Table 1 Suggested source concepts

Target concepts
LIFE
DEATH
TIME/MOMENT IN TIME
FUTURE
PAST
CHANGE
PROGRESS/EVOLUTION/DEVELOPMENT
SUCCESS/ACCOMPLISHMENT
CAREER
FEELINGS/EMOTIONS
ATTITUDES/VIEWS
MIND
IDEAS
KNOWLEDGE
PROBLEM
TASK/DUTY/RESPONSIBILITY
VALUE
WELL-BEING
SOCIAL/ECONOMIC/POLITICAL SYSTEM
RELATIONSHIP

Table 2 Suggested target concepts

their approach to distinguishing basic senses from metaphorical ones. We modify and extend the procedure to identify source-target domain mappings by comparing

the contexts in which a word appears in its basic and metaphorical senses. Besides assigning labels to metaphorical associations, this stage of the procedure then feeds back into the metaphor identification process and acts as an additional constraint on metaphoricity.

Since the experiments involving multiple annotators focus on metaphors expressed by a verb, the annotation procedure and guidelines, although in principle suitable for the analysis of all parts of speech, were tailored to verb metaphors. The procedure used as part of annotation guidelines is presented below.

1. For each verb establish its meaning in context and try to imagine a more basic meaning of this verb in other contexts. As defined in the framework of MIP [47] basic meanings are normally:
   - more concrete;
   - related to bodily action;
   - more precise (as opposed to vague);
   - historically older.
2. If you can establish a basic meaning that is distinct from the meaning of the verb in this context, the verb is likely to be used metaphorically. Try to identify a mapping between the source domain (where the basic meaning comes from) and the target domain (the concepts forming the context of the verb in front of you) using the provided lists of source and target categories. Record the mapping. If you fail to identify a mapping, reconsider whether the sense is really metaphorical in this context.

The following example illustrates how the procedure operates in practice.

(11) If he asked her to post a letter or buy some razor blades from the chemist, she was transported with pleasure.

In this sentence one needs to annotate the four verbs that are underlined.

- The first 3 verbs are used in their basic sense, i.e. literally (*ask* in the context of "a person asking another person a question or a favour"; *post* in the context of "a person posting/sending a letter by post"; *buy* in the sense of "making a purchase"). Thus they are tagged as literal.
- The verb *transport*, however, in its basic sense is used in the context of "goods being transported/carried somewhere by a vehicle". The context in this sentence involves "a person being transported by a feeling", which contrasts with the basic sense in that the agent of *transporting* is an EMOTION (the target concept) as opposed to a VEHICLE (the source concept). Thus one can infer that the use of *transport* in this sentence is metaphorical and the associated interconceptual mapping is EMOTIONS – VEHICLES.

In our experiments, the annotators were asked to imagine the contexts in which the verb has a more basic meaning, as opposed to choosing from a predefined set of contexts or using a dictionary. Provided that the basic meaning satisfies the definition and the properties given in the annotation guidelines, the senses they could

select were unrestricted. This distinguishes our procedure from the work of [47], that had a stronger reliance on dictionary definitions for this purpose. We believe that a procedure allowing for some flexibility in combination with clear definitions is better suited for the analysis of metaphorical meanings than a strictly-regulated dictionary-based procedure, since the metaphorical meanings themselves are a dynamic and flexible phenomenon. While a dictionary-based analysis is likely to result in an increased inter-annotator agreement, there is a risk that it may leave a number of word senses and metaphorical mappings unaccounted for. According to previous studies [34], the inclusion of metaphorical senses in dictionaries and lexical resources is often unsystematic: some conventional metaphorical senses are included in the dictionaries, while others are omitted.

5 Annotation reliability study

After an annotation scheme has been developed its reliability needs to be verified. Reliability of a scheme can be assessed by comparing annotations carried out by multiple annotators independently [26]. This section describes an experiment where the same small portion of the metaphor corpus was annotated by several participants.

5.1 Data

A text sample from the BNC (text ID: ACA) was selected for the reliability study. Since the focus of the study is on single-word metaphors expressed by a verb, the first part of the annotation task can be viewed as verb classification according to whether the verbs are used metaphorically or literally. However, some verbs inherently have a weak potential, or no potential at all, to be used metaphorically, and as such the study is not concerned with them. The following verb classes were excluded: (1) auxiliary verbs; (2) modal verbs; (3) aspectual verbs (e.g. *begin, start, finish*); and (4) light verbs (e.g. *take, give, put, get, make*).

5.2 Annotation experiment

Subjects Three independent volunteer annotators participated in the experiment. They were native speakers of English and held a graduate degree in linguistics or computer science. However, they were naive to the specific purposes of the study and the claims of the CMT.

Material and Task The subjects were given the same text from the BNC which was a social science essay. The text contained 142 verbs to annotate, which were

underlined. They were asked to (1) classify verbs as metaphorical or literal, and (2) identify the source-target domain mappings for the verbs they marked as metaphorical. They received two lists of suggested categories describing source and target concepts, and were asked to select a pair of categories from the two lists that best described the metaphorical mapping. Along with this they were allowed to introduce new categories if they felt none of the given categories expressed the mapping well enough. The annotation was done electronically using colour highlighting and inserting category labels in Microsoft Word.

Guidelines and Training The annotators received written instructions (2 pages, corresponding to the guidelines described in the previous section) and were asked to do a small annotation exercise (2 sentences: 1 example sentence and 1 sentence to annotate, containing 8 verbs in total). The goal of the exercise was to ensure they were at ease with the annotation format.

5.3 Interannotator agreement

Semantic annotations involve interpretation on the part of the participant and are thus inherently subjective. It is therefore essential to report *interannotator agreement*, that quantifies the similarity of the annotations produced by different annotators. We evaluated reliability of the proposed annotation scheme by assessing interannotator agreement in terms of κ statistic [54] on both tasks separately.

Kappa statistic As opposed to simple percentage of identically tagged instances, κ measures agreement by factoring out the degree of agreement expected by chance. It is calculated as follows:

$$\kappa = \frac{P(A) - P(E)}{1 - P(E)}, \tag{1}$$

where $P(A)$ is the proportion of times that k annotators agree and $P(E)$ the proportion of times one would expect them to agree by chance. If there is perfect agreement among the annotators, then $\kappa = 1$, whereas if there is no agreement besides what is expected by chance, then $\kappa = 0$. κ is negative when annotators agree less than expected by chance. The amount of agreement one would expect to happen by chance depends on the number and distribution of the categories used for annotation.[3] The values of κ are traditionally interpreted using the following scale [31]: 0 - 0.20 indicates slight agreement, 0.21 - 0.40 fair agreement, 0.41 - 0.60 moderate agreement, 0.61 - 0.80 substantial agreement, 0.81 - 1 means the agreement is almost perfect.

Results The number of metaphors and their conceptual mappings as annotated by the participants are shown in Table 3. The average proportion of the cases where a conceptual metaphor could be annotated for a given linguistic metaphor (across

[3] For more details on how chance agreement is calculated see [54].

the three annotators) was 95%, whereas that using the categories from the provided lists was 82%.

The reliability of the scheme was first measured for the task of metaphor identification and then for the assignment of interconceptual mappings. The identification of metaphorical verbs yielded a reliability of $\kappa = 0.64$ $(n = 2; N = 142; k = 3)$, where n stands for the number of categories, N for the number of instances annotated and k for the number of annotators. This level of agreement is considered substantial.

The measurement of the agreement in the second task appeared less straightforward. It was complicated by the fact that each annotator only assigned conceptual mappings to a set of verbs that in their judgement were metaphorical. These sets were not identical for all annotators. Thus, the agreement on the assignment of source and target domain categories was calculated only using the instances that all annotators considered to be metaphorical. This yielded a total of 30 conceptual mappings to compare.

One of the annotators (C) found the provided categories insufficient. Although trying to use them where possible, he nonetheless had to introduce a large number of categories of his own to match his intuitions, which generally suggests the insufficiency of MML. In addition, he did not assign any mapping for seven metaphorical expressions. Both of these issues complicated the comparison of his annotation to those of the other annotators. Thus, his labelling of the mappings was excluded from the calculation of kappa statistic for agreement on conceptual metaphor annotation. However, his data was qualitatively analysed along with the rest.

Annotator	Metaphors	Annotated mappings	Target from list	Source from list
A	53	53	52	52
B	39	39	39	37
C	58	51	42	25

Table 3 Differences in annotations

The resulting overall agreement on the assignment of conceptual metaphor was thus $\kappa = 0.57$ $(n = 26; N = 60; k = 2)$, whereby the agreement was stronger on the choice of the target categories ($\kappa = 0.60$ $(n = 14; N = 30; k = 2)$) than the source categories ($\kappa = 0.54$ $(n = 12; N = 30; k = 2)$).

Analysis of annotations Analysing cases of disagreement during metaphor identification suggests that the main source of disagreement was the conventionality of some metaphorical uses. These include expressions whose metaphorical etymology can be clearly traced, but the senses are lexicalised (e.g. "*fall* silent", "the end is *coming*") and thus perceived by some annotators as literal.

According to the annotators' informal feedback on the experiment, they found the task of identifying linguistic metaphor relatively straightforward, whereas the task of assigning the respective conceptual metaphor appeared more difficult. The analysis of annotations has shown that one of the sources of disagreement in the latter task was the presence of partially overlapping categories in the target concept

list. For example, the categories of PROGRESS and SUCCESS, or VIEWS, IDEAS and METHODS were often confused. This level of granularity was chosen following the Master Metaphor List. However, the annotated data suggests that, for the purpose of annotation of conceptual mappings, such categories may be joined into more general categories without significant information loss (e.g. VIEWS, IDEAS and METHODS can be covered by a single category IDEAS). This would increase mutual exclusivity of categories and thus lead to a more consistent annotation. Based on the observations in the data and the annotators' feedback, the source and target lists were refined to ensure no or minimal overlap between the categories, while maximally preserving their informativeness. As a post-hoc experiment, the labels in the annotations were mapped to this new set of categories and the annotations were compared again. The agreement rose to $\kappa = 0.61$ $(n = 23; N = 60; k = 2)$, as expected.

Further examples of similarities and differences in the annotations are given in Figure 2. As the examples illustrate, the annotators tend to agree on whether a verb is used metaphorically or literally (with the exception of the verb *catch* tagged as literal by Annotator B). Their choices of source and target domain categories, however, vary. The annotators often choose the same target domain, although they refer to it by different (overlapping) labels, e.g. IDEA/THOUGHT/VIEW or TIME/MOMENT IN TIME. Annotator C introduced a more general category PERCEPTION, rather than using the more specific category VISION provided in the list, or DISEASE instead of the suggested category INFECTION. Thus they tend to choose categories that are intuitively related and the variation of the target domain labels is rather due to the granularity of categories used. In contrast, the choice of the source domain labels exhibits more conceptual variation. Annotator A tends to assign a general category PHYSICAL OBJECT to all instances appearing within the context related to physical activity, whereas Annotator B opts for finer-grained categories, as well as conceptualising the context in terms of events and actions rather than objects. These observations suggest that, although the annotators may share some of the intuitions with respect to conceptual metaphor, the explicit labelling of the latter in text is a challenging task. Furthermore, the across-annotator variability can be seen as problematic for the CMT, as it is inconsistent with the idea that there are fixed mappings between conceptual domains, with knowledge in one domain being generally understood in terms of knowledge in another.

6 Corpus data analysis

In order to create a dataset for experimentation, as well as to perform a more comprehensive data analysis, a single annotator annotated a larger corpus using the above procedure. The corpus contains 761 sentences and 13,642 words. The text used for the reliability study constituted a part of the corpus and the same set of source and target categories was employed. This allowed to measure the agreement with the external annotators. The agreement on the identification of linguistic meta-

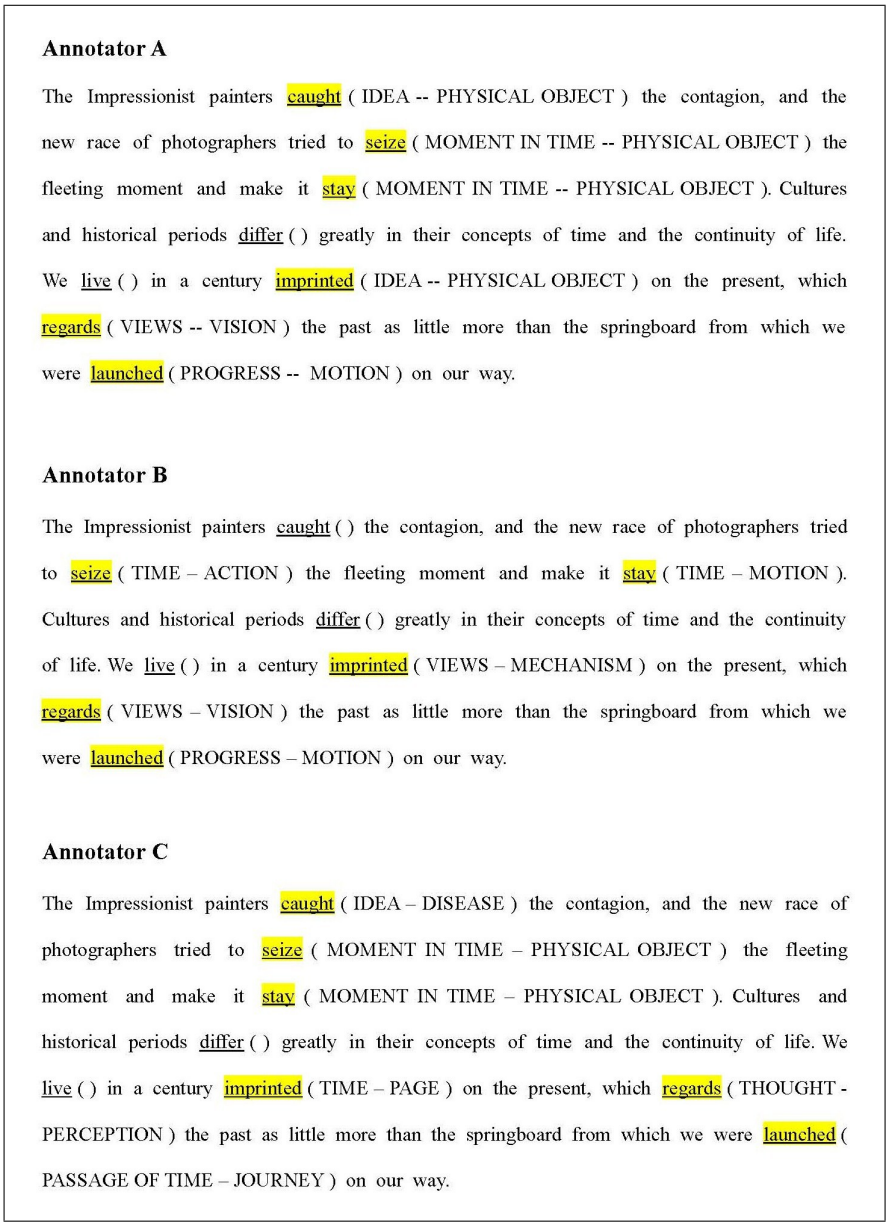

Fig. 2 Example of similarities and differences in annotation

phor was $\kappa = 0.62$ $(n = 2; N = 142; k = 4)$, whereas that on the choice of source and target domain categories reached $\kappa = 0.58$ $(n = 22; N = 56; k = 3)$.

As an additional experiment, we also annotated nouns, adjectives and adverbs in the corpus as metaphorical or literal using the same procedure. This was done in order to investigate how metaphor can be expressed by other word classes, to gather metaphor statistics across a wider range of syntactic constructions and to estimate the relative proportion of verbal metaphors across genres (the study by [6] only concerned metaphor in educational discourse). In what follows we will describe statistics of the resulting corpus and attempt to identify common traps in the annotation of source-target domain mappings in real-world text.

6.1 Metaphor statistics across genres

Metaphor frequency was calculated as the number of metaphors relative to the number of sentences in the text. The results presented in Table 4 indicate that metaphor is overall an extremely frequent phenomenon - it appears on average in every third sentence. An interesting finding is that fiction texts seem to contain fewer metaphors than other genres. However, it should be noted that the frequency metric used is biased towards genres with longer sentences, and fiction texts contain some dialogues consisting of short phrases. In addition, the dialogues themselves tend to contain mainly literal language, as opposed to author's descriptions where metaphors are more frequent. Overall, therefore, fiction contains relatively fewer metaphorical expressions than other genres.

Text	ID	Genre	Sent.	Words	Met–rs	Met./Sent.	Verb m.
Hand in Glove, Goddard	G0N	Literature	335	3927	41	0.12	30
After Gorbachev, White	FYT	Politics	45	1384	23	0.51	17
Today newspaper	CEK	News	116	2086	48	0.41	30
Tortoise by Candlelight, Bawden	HH9	Literature	79	1366	12	0.15	10
The Masks of Death, Cecil	ACA	Sociology	60	1566	70	1.17	42
Radio broadcast (current affairs)	HM5	Speech	58	1828	10	0.17	7
Language and Literature journal	J85	Article	68	1485	37	0.54	28
Total			761	13642	241	0.32	164

Table 4 Corpus statistics for metaphor

The last column of Table 4 shows the proportion of verb metaphors across genres. The distribution of their frequency over genres appears similar to that of other part of speech classes. However, it should be noted that metaphors expressed by a verb are by a large margin the most frequent type and constitute 68% of all metaphorical expressions in the corpus.

Frequency	Source concepts
0.23	MOTION
0.13	VISION/SEEING
0.13	LIVING BEING
0.13	GROWTH/RISE
0.07	SPEED
0.03	DIRECTIONALITY: e.g. UP/DOWN
0.03	BASIS/PLATFORM
0.03	LOCATION
0.03	DISTANCE
0.03	MACHINE/MECHANISM
0.03	PHYSICAL OBJECT
...	

Table 5 Distribution of source concepts

Frequency	Target concepts
0.27	ATTITUDES/VIEWS
0.13	CHANGE
0.12	TIME/MOMENT IN TIME
0.12	PROGRESS/EVOLUTION/DEVELOPMENT
0.05	BEHAVIOUR
0.05	SUCCESS/ACCOMPLISHMENT
0.05	FUTURE
0.05	CAREER
0.03	SOCIAL/ECONOMIC/POLITICAL SYSTEM
0.03	IDEAS
0.03	METHODS
0.03	KNOWLEDGE
0.02	DEATH
0.02	PAST

Table 6 Distribution of target concepts

6.2 Mappings statistics

It is also interesting to look at the distributions of the source and target categories in the text annotated by the three annotators, shown in Tables 5 and 6 respectively. The topic of the text (in this case sociology) has an evident influence on the kind of mappings that can be observed in this text.

The most frequent source domain of MOTION was mainly mapped onto the target concepts of CHANGE, PROGRESS, CAREER and SUCCESS. TIME was generally associated with DISTANCE, and the MOMENT IN TIME category with LOCATION. VIEWS and IDEAS were viewed as either LIVING BEINGS or PHYSICAL OBJECTS. A large proportion of the mappings identified match those exemplified in the Master Metaphor List, but some of the mappings suggested by the annotators are novel (for example, EMPHASIS IS A PHYSICAL FORCE, SITUATION IS A PICTURE, etc).

6.3 Interaction of metaphor and metonymy

An interesting issue observed in the data is the combination of metaphor and metonymy within a phrase. Consider the following example:

(12) We live in a century *imprinted* on the present, which *regards* the past as little more than the springboard from which we were *launched* on our way. (BNC: ACA)

In this sentence the verbs *imprint*, *regard* and *launch* are used metaphorically according to all annotators. However, the noun *present* can be interpreted as a general metonymy referring to the people who live in the present, rather than the time period. In the latter case, the verb *regard* would receive a different, more conventional interpretation. This in turn is likely to affect the annotation of the corresponding conceptual metaphor and may even result in *regard* being tagged as literally used.

7 Challenges in metaphor annotation and lessons learned

The current study also revealed a number of difficulties in the annotation of source-target domain mappings in real-world text. This section discusses the main challenges in metaphor annotation and the lessons learned from the study.

7.1 Level of generality and relations between the mappings

One of the major steps in the design of the annotation scheme for conceptual metaphor is the construction of the inventory of categories that generalise across many metaphorical expressions. However, given a set of examples, it is often unclear at which level of generality the source and target categories should stand. Consider the following sentence:

(13) Sons aspired to *follow* ((CAREER or LIFE) IS A (PATH or JOURNEY)) in their fathers' trades or professions.

Here the verb *follow* is used metaphorically; the best generalisations for both source and target domains are, however, not obvious. This metaphor can be characterised by a more precise mapping of CAREER IS A PATH, as well as the general one of LIFE IS A JOURNEY, or a mix of the two. These two mappings are related, however, the nature of this relationship is not entirely clear. [38] discusses hierarchical organisation of conceptual metaphors and models it in terms of subsumption. [30] point out cases of entailment relations between mappings, e.g. the metaphor TIME IS MONEY entails TIME IS A VALUABLE COMMODITY or TIME IS A LIMITED RESOURCE. This entailment is based on the fact that the source concepts in the latter mappings are properties of MONEY. However, the more general

metaphor LIFE IS A JOURNEY does not strictly entail or subsume the metaphor CAREER IS A PATH. CAREER is not necessarily a property of LIFE, but is part of one possible life scenario, in which career is present and is an important variable. [10] view metaphor in terms of such discrete scenarios within the domains, rather than in terms of continuous domains themselves. Originating in the source domain, the scenarios can then be applied to reason about the target domain. Thus certain scenarios from the domain of JOURNEY can be projected onto the domain of LIFE, for example, describing the concept of CAREER through that of a PATH. Viewing source and target domains as continuous rather than discrete concepts is in line with the insights one may gain from our study. This in turn suggests that, although cross-domain metaphorical mappings clearly exist, it may not be optimal to describe them using discrete natural-language labels. It is likely that a different representation is needed, one that allows us to capture the vagueness, fuzziness and variability, that are inherent in the use of metaphor. Even though the ability of Lakoff's CMT to explain metaphorical language in principle is evident from our results, the question of labelling source and target domains remains open.

7.2 Chains of mappings

Another challenge revealed by our study is that in some cases chains of mappings are necessary to explain a metaphorical expression. Consider the following example:

(14) The Impressionist painters *caught the contagion* [..] (BNC: ACA)

In this sentence the phrase *caught the contagion* is used metaphorically. The interpretation of this metaphor triggers two conceptual mappings, namely IDEAS/VIEWS – INFECTION and INFECTION – PHYSICAL OBJECT. This chain-like association structure intuitively seems natural to a human. At the same time, though, it brings additional complexity into the annotation process, since the number of associations involved may vary. However, it should be noted that the cases where chains of mappings are necessary to explain a metaphorical expression are rare, and only three examples of this phenomenon were found in the corpus.

7.3 Annotation scheme: Lessons learned

Besides gaining further insights into CMT, our metaphor annotation experiment also sheds light on a number of issues concerning the annotation process itself.

When designing a metaphor annotation scheme one faces a choice of either employing a dictionary or relying on the annotators' imagination in order to compare the possible literal and metaphorical contexts for the given word (verb). While previous work relied on dictionary definitions for this purpose, our annotation scheme is more flexible, allowing the annotators to imagine the literal context of every word.

While the former is likely to increase the interannotator agreement, the latter is more suitable to capture the dynamic properties of metaphor and the freedom of interpretation associated with them. Our results confirm that the subjects were able annotate metaphorical expressions without the use of a dictionary with substantial reliability.

The main source of disagreement between the annotators was the conventionality of some metaphorical expressions. Highly conventional metaphors were tagged as literal by some annotators, and as metaphorical by others. Since we already know that the literal-metaphorical distinction is not clear-cut, one possible solution would be to introduce metaphor annotation on a gradual scale. Such scale can be defined, for example, from *strongly literal*, to *somewhat metaphorical* (highly conventional, but exhibiting some metaphorical properties), to *strongly metaphorical* expressions. Introducing a scale may make the task easier for annotators, who in the current scheme were confined to taking binary decisions about inherently fuzzy categories. Annotation on a gradual scale would also better reflect the nature of the phenomenon, and situate metaphorical language on a continuum, as pointed out by [15]. It would thus allow us to study the role of conventionality in metaphor interpretation, and potentially yield new, informative insights about how metaphor should be modelled.

Annotations of conceptual metaphor had two main sources of variation: (1) the annotators introduced a number of their own unique concepts; and (2) the annotated source and target concepts differed in their level of generality. The fact that the annotators needed to bring in concepts from new domains and topics in order to describe their intuitions about specific conceptual metaphors suggests that the coverage of the provided source and target domain lists was insufficient. This is mainly due to how the list was compiled, rather than confirming or refuting the fundamental claims of CMT. However, the fact that some of the differences between the annotations stemmed from the different level of generality of the chosen concepts in the conceptual hierarchy, suggests that it is hard to pre-define an exhaustive list of source and target domains labels even in principle. And thus if metaphor annotation is carried out in terms of explicit natural language labels, it should rely on the annotators selecting their own categories based on their analysis of linguistic contexts, rather than on a predefined set.

It should also be noted that joining categories together and making them more general and mutually exclusive increases the annotation reliability, and possibly makes the task easier. However, the analysis of the data suggests that it is very hard (and potentially impossible) to annotate conceptual metaphor in terms of mutually exclusive categories without loss of information.

8 What kind of metaphor annotation does NLP need?

The problem of metaphor modelling is steadily gaining interest in NLP. However, there is still no single task definition or shared dataset against which the systems can be evaluated. This makes it hard to directly compare the systems and draw con-

clusions about the benefits and drawbacks of particular approaches. The annotation scheme and experiment we presented is a step towards creating a general framework for metaphor annotation and a dataset for system evaluation. However, much like other such schemes, it has been primarily motivated by the linguistic considerations and the desire to verify some of the claims of CMT. But what kind of metaphor annotation does NLP need and to what extent does our scheme satisfy these criteria? And how can we integrate the lessons learned from this experiment into the design and evaluation of metaphor processing systems?

One of the primary questions is whether NLP needs a model of conceptual metaphor, or is processing linguistic metaphor enough? Strictly speaking, NLP systems need to be able to interpret textual data, and thus they need to address linguistic metaphor in the first place. In order to perform the interpretation of linguistic metaphors, the system may or may not use a conceptual metaphor representation. On one hand, a suitable representation of conceptual metaphor may inform the system's decisions regarding linguistic metaphor, possibly increasing the system's accuracy. However, it is unlikely to form an important part of system functionality on its own. Thus metaphor annotation does not necessarily need to be concerned with assigning source and target domain labels, and may focus on linguistic metaphor alone.

If one decided to build a computational model of conceptual metaphor nonetheless, it is important to consider how conceptual metaphor should be represented within the system. Our study demonstrated that the assignment of labels to source and target domains is a challenging task. This suggests that it is preferable to model source and target domain mappings implicitly within the system, rather than setting a goal of assigning explicit domain labels automatically. Implicit modeling of conceptual metaphor has been successfully exploited by some metaphor processing systems [53, 52], while others assigned explicit labels in the form of clusters or WordNet synsets and then manually mapped them to labels from the Master Metaphor List [41]. This work has shown that mapping the system-learned representations of conceptual metaphor to any labels in the manually-annotated data is a non-trivial task, an it is not clear if the annotations would provide an objective feedback to the system.

NLP systems are thus mainly concerned with the evaluation of linguistic metaphor, and require primarily corpus data annotated for linguistic metaphor. Our scheme provides one way of creating this kind of data. However, one issue that remains open is metaphor conventionality, and it is not yet clear where in the metaphorical–literal continuum the system should draw the line between what it considers metaphorical and what it considers literal. The answer to this question most likely depends on the NLP application in mind. However, generally speaking, real-world NLP applications are unlikely to be concerned with historical aspects of metaphor, but rather with the identification of figurative language that needs to be interpreted differently from literal language. We, therefore, suggest that NLP applications do not necessarily need to address highly conventional metaphors that can be interpreted using standard word sense disambiguation techniques, but rather would benefit from the identification of less conventional and more creative language. Metaphor annotation efforts should thus bear this distinction in mind.

9 Conclusion

Besides making our thoughts more vivid and filling our communication with richer imagery, metaphors also play an important structural role in our cognition [30]. This chapter described a flexible scheme for annotation of metaphorical associations in arbitrary text and an annotation study that allowed us to gain further insight into the inner workings of this important and fascinating phenomenon. The annotation scheme was designed with open-domain metaphor annotation in mind, enabling the study of CMT in real-world data. Metaphorical mappings are annotated by explicit context comparison, and source and target domain labels are assigned to the contexts independently, rather than in the form of a preconstructed mapping.

Our annotation experiment has shown that metaphor is a highly frequent phenomenon, which makes its thorough investigation indispensable for theoretical and applied, cognitive and computational study of language. Another important finding is that a large proportion of linguistic metaphors (68%) are represented by verbs, which provides a post-hoc justification for our choice of verbal constructions for this study.

We then investigated how conceptual metaphor manifests itself in language. Although the annotators reach some overall agreement on the annotation of interconceptual mappings, they experienced a number of difficulties. The greatest of them was the problem of finding the right level of abstraction for the domain categories. The difficulties in category assignment suggest that it is hard to consistently assign explicit labels to source and target domains, even though the interconceptual associations exist in some sense and are intuitive to humans. Awareness of these issues can potentially feed back to CMT or other theoretical accounts of metaphor. Such problems also need to be taken into account when designing a cognitive or computational model of metaphor that relies on CMT. A certain degree of vagueness and freedom in interpretation is one of the purposes of metaphorical language, which makes metaphor a challenging task for computational modelling. A computational model needs to operate over a well-defined set of categories (either manually listed or automatically learned) and their consistency and coverage would then play a crucial role in how well the model can account for real-world data. We believe that the results of our annotation study, despite indicating that the metaphorical mappings themselves are intuitive to humans (i.e. they can be annotated in arbitrary text), still show that the a predefined set of categories, such as those widely discussed in linguistic literature on CMT, may not be sufficient or even suitable for a computational model. And despite the validity of the main principles of CMT as a linguistic theory, it is not straightforward to port it to computational modelling of metaphor and a more flexible, and potentially data-driven, representation of source and target domain categories is needed for this purpose. A data-driven representation would also be better suited to account for the freedom in interpretation of metaphor, as it can be dynamically learned from the data.

As an alternative to explicit source and target domain labels, source and target domains could be represented, for example, as classes, or clusters, of related concepts, optimized to capture the majority of metaphorical instances. For example, in (13)

the concept of CAREER can be clustered together with the concept of LIFE, and the resulting cluster can then represent the target domain. Such clusters of concepts may be learned empirically from linguistic data, as shown by [53]. The individual clusters can then be organised into a network, where the links between the clusters represent metaphorical associations. A detailed description, design and verification of such a model are, however, left for future work.

Finally, the corpus presented here provides a new dataset for linguistic, computational and cognitive research on metaphor. Further empirical studies of the interconceptual mappings in real-world linguistic data may shed light on the way metaphorical associations govern our reasoning processes and organize our conceptual system, in terms of which we think, communicate, create and act.

References

1. Agerri, R., J.A. Barnden, M.G. Lee, and A.M. Wallington. 2007. Metaphor, inference and domain-independent mappings. In *Proceedings of RANLP-2007*, pages 17–23, Borovets, Bulgaria.
2. Barnden, J.A. and M.G. Lee. 2002. An artificial intelligence approach to metaphor understanding. *Theoria et Historia Scientiarum*, 6(1):399–412.
3. Black, M. 1962. *Models and Metaphors*. Cornell University Press.
4. Bowdle, Brian F. and D. Gentner. 2005. The career of metaphor. *Psychological Review*, 112:193–216.
5. Burnard, L. 2007. *Reference Guide for the British National Corpus (XML Edition)*.
6. Cameron, L. 2003. *Metaphor in Educational Discourse*. Continuum, London.
7. Charniak, E., D. Blaheta, N. Ge, K. Hall, J. Hale, and M. Johnson. 2000. *BLLIP 1987-89 WSJ Corpus Release 1*. Linguistic Data Consortium, Philadelphia.
8. Chung, S. F., K. Ahrens, C. R. Huang. 2005. Source Domains as Concept Domains in Metaphorical Expressions. *International Journal of Computational Linguistics and Chinese Language Processing*, 10(4): 553-570.
9. Deignan, A. 2006. The grammar of linguistic metaphors. In A. Stefanowitsch and S. T. Gries, editors, *Corpus-Based Approaches to Metaphor and Metonymy*, Berlin. Mouton de Gruyter.
10. Fauconnier, G. and M. Turner. 2002. *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books.
11. Feldman, J. and S. Narayanan. 2004. Embodied meaning in a neural theory of language. *Brain and Language*, 89(2):385–392.
12. Feldman, J. A. 2006. *From Molecule to Metaphor: A Neural Theory of Language*. The MIT Press.
13. Gentner, D. 1983. Structure mapping: A theoretical framework for analogy. *Cognitive Science*, 7:155–170.
14. Gentner, D., Mutsumi Imai, and Lera Boroditsky. 2002. As time goes by: Evidence for two systems in processing space-time metaphors. *Language and Cognitive Processes*, 47:537–565.
15. Gibbs, R. 1984. Literal meaning and psychological theory. *Cognitive Science*, 8:275–304.
16. Gibbs, R. and M. Tendahl. 2006. Cognitive effort and effects in metaphor comprehension: Relevance theory and psycholinguistics. *Mind & Language*, 21:379–403.
17. Glucksberg, S. 2003. The psycholinguistics of metaphor. *Trends in Cognitive Science*, 7:92–96.
18. Goatly, A. 1997. *The Language of Metaphors*. Routledge, London.
19. Gong, S. P., K. Ahrens and C. R. Huang. 2008. Chinese Word Sketch and Mapping Principles: A Corpus-Based Study of Conceptual Metaphors Using the BUILDING Source Domain. *International Journal of Computer Processing of Oriental Languages*, 21(2):3–17.

20. Grady, J. 1997. Foundations of meaning: primary metaphors and primary scenes. Technical report, PhD thesis, University of California at Berkeley.
21. Hardie, A., Koller, V., Rayson, P. and Semino, E. 2007. Exploiting a Semantic Annotation Tool for Metaphor Analysis. *Proceedings of the Corpus Linguistics Conference*, Birmingham, UK.
22. Haskell, R. E. 2002. Cognitive science and the origin of lexical metaphor. *Theoria et Historia Scientiarum*, 6(1):291–331.
23. Izwaini, S. 2003. Corpus-based Study of Metaphor in Information Technology. In *Proceedings of the Workshop on Corpus-based Approaches to Figurative Language, Corpus Linguistics 2003*, Lancaster, 27 March 2003.
24. Keysar, B., Y. Shen, S. Glucksberg, and W.S. Horton. 2000. Conventional language: How metaphorical is it? *Journal of Memory and Language*, 43:576–593.
25. Koivisto-Alanko, P. and H. Tissari. 2006. Sense and sensibility: Rational thought versus emotion in metaphorical language. In A. Stefanowitsch and S. T. Gries, editors, *Corpus-Based Approaches to Metaphor and Metonymy*, Berlin. Mouton de Gruyter.
26. Krippendorff, K. 1980. *Content Analysis*. SAGE Publications, Beverly Hills, CA.
27. Krishnakumaran, S. and X. Zhu. 2007. Hunting elusive metaphors using lexical resources. In *Proceedings of the Workshop on Computational Approaches to Figurative Language*, pages 13–20, Rochester, NY.
28. Lakoff, G. 1992. The contemporary theory of metaphor. In A. Ortony, editor, *Metaphor and Thought (2nd ed)*, pages 202–251, Cambridge, U.K. Cambridge University Press.
29. Lakoff, G., J. Espenson, and A. Schwartz. 1991. The master metaphor list. Technical report, University of California at Berkeley.
30. Lakoff, G. and M. Johnson. 1980. *Metaphors We Live By*. University of Chicago Press, Chicago.
31. Landis, J. and G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.
32. Lönneker, B. 2004. Lexical databases as resources for linguistic creativity: Focus on metaphor. In *Proceedings of the LREC 2004 Workshop on Language Resources for Linguistic Creativity*, pages 9–16, Lisbon, Portugal.
33. Lönneker-Rodman, B. 2008. The Hamburg Metaphor Database Project. Issues in Resource Creation *Language Resources and Evaluation*, 42: 293–318.
34. Lönneker, B and C. Eilts. 2004. A Current Resource and Future Perspectives for Enriching WordNets with Metaphor Information. In *Proceedings of the Second International WordNet Conference (GWC 2004)*, pages 157–162, Brno, Czech Republic, 2004.
35. Low, G., Z. Todd, A. Deignan and L. Cameron. 2010. *Researching and Applying Metaphor in the Real World*. John Benjamins, Amsterdam/Philadelphia.
36. Lu, L. and K. Ahrens. 2008. Ideological Influences on BUILDING metaphors in Taiwanese Presidential Speeches. *Discourse and Society*, 19(3):383–408.
37. Martin, J. H. 1988. Representing regularities in the metaphoric lexicon. In *Proceedings of the 12th conference on Computational linguistics*, pages 396–401.
38. Martin, J. H. 1990. *A Computational Model of Metaphor Interpretation*. Academic Press Professional, Inc., San Diego, CA, USA.
39. Martin, J. H. 1994. Metabank: A knowledge-base of metaphoric language conventions. *Computational Intelligence*, 10:134–149.
40. Martin, J. H. 2006. A corpus-based analysis of context effects on metaphor comprehension. In A. Stefanowitsch and S. T. Gries, editors, *Corpus-Based Approaches to Metaphor and Metonymy*, Berlin. Mouton de Gruyter.
41. Mason, Z. J. 2004. Cormet: a computational, corpus-based conventional metaphor extraction system. *Computational Linguistics*, 30(1):23–44.
42. McGlone, Matthew S. 2007. What is the explanatory value of a conceptual metaphor? *Language & Communication*, 27:109–126.
43. Murphy, G. L. 1996. On metaphoric representation. *Cognition*, 60:173–204.
44. Narayanan, S. 1997. Knowledge-based Action Representations for Metaphor and Aspect (KARMA). Technical report, PhD thesis, University of California at Berkeley.

45. Narayanan, S. 1999. Moving right along: A computational model of metaphoric reasoning about events. In *Proceedings of AAAI 99)*, pages 121–128, Orlando, Florida.
46. Pinker, S. 2007. *The Stuff of Thought: Language as a Window into Human Nature*. Viking Adult, USA, September.
47. Pragglejaz Group. 2007. MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22:1–39.
48. Reddy, M. 1978. The conduit metaphor: A case of frame conflict in our language about language. In A. Ortony, editor, *Metaphor and Thought (2nd ed)*, pages 164–201, Cambridge, U.K. Cambridge University Press.
49. Shalizi, C. R. 2003. Analogy and Metaphor.
50. Shutova, E. 2010. Models of Metaphor in NLP. In *Proceedings of ACL 2010*, Uppsala, Sweden.
51. Shutova, E., S. Teufel. 2010. Metaphor Corpus Annotated for Source-Target Domain Mappings. In *Proceedings of LREC 2010*, Valletta, Malta.
52. Shutova, E. and L. Sun. 2013. Unsupervised metaphor identification using hierarchical graph factorization clustering. In *Proceedings of NAACL 2013*, Atlanta, GA, USA.
53. Shutova, E., L. Sun, and A. Korhonen. 2010. Metaphor identification using verb and noun clustering. In *Proceedings of Coling 2010*, Beijing, China.
54. Siegel, S. and N. J. Castellan. 1988. *Nonparametric statistics for the behavioral sciences*. McGraw-Hill Book Company, New York, USA.
55. Skorczynska Sznajder, H. and J. Pique-Angordans. 2004. A corpus-based description of metaphorical marking patterns inscientific and popular business discourse. In *Proceedings of European Research Conference on Mind,Language and Metaphor (Euresco Conference)*, Granada, Spain.
56. Steen, G. J., A. G. Dorst, J. B. Herrmann, A. A. Kaal, T. Krennmayr, and T. Pasma. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*. John Benjamins, Amsterdam/Philadelphia.
57. Wallington, A. M., J. A. Barnden, P. Buchlovsky, L. Fellows, and S. R. Glasbey. 2003. Metaphor Annotation: A Systematic Study. Technical report, School of Computer Science, The University of Birmingham.
58. Wikberg, K. 2006. The Role of Corpus Studies in Metaphor Research. *2006 Stockholm Metaphor Festival, eds. N.-L. Johannesson and D.C. Minugh.*
59. Wilks, Y. 1975. A preferential pattern-seeking semantics for natural language inference. *Artificial Intelligence*, 6:53–74.