The Curious Case of Metonymic Verbs:
A Distributional Characterization

Jason Utt[1], Alessandro Lenci[2], Sebastian Padó[3], Alessandra Zarcone[1]

[1]Universität Stuttgart, [2]Universität Heidelberg, [3]Università di Pisa

[1]uttjn,zarconaa@ims.uni-stuttgart.de,
[2]alessandro.lenci@ling.unipi.it, [3]pado@cl.uni-heidelberg.de

Abstract

Logical metonymy combines an event-selecting verb with an entity-denoting noun (e.g., *The writer began the novel*), triggering a covert event interpretation (e.g., *reading*, *writing*). Experimental investigations of logical metonymy must assume a binary distinction between metonymic (i.e. event-selecting) verbs and non-metonymic verbs to establish a control condition. However, this binary distinction (whether a verb is metonymic or not) is mostly made on intuitive grounds, which introduces a potential confounding factor.

We describe a corpus-based approach which characterizes verbs in terms of their behavior at the syntax-semantics interface. The model assesses the extent to which transitive verbs prefer event-denoting objects over entity-denoting objects. We then test this "eventhood" measure on psycholinguistic datasets, showing that it can distinguish not only metonymic from non-metonymic verbs, but that it can also capture more fine-grained distinctions among different classes of metonymic verbs, putting such distinctions into a new graded perspective.

1 Motivation

Logical metonymy, an instance of *enriched composition* (Jackendoff, 1997), consists of a combination of an event-selecting metonymic verb and an entity-denoting direct (e.g., *The writer began the novel*).[1] Its interpretation involves the recovery of a *covert event* (*reading*, *writing*). Metonymy interpretation is generally explained in terms of a type clash between the verb's selectional restrictions and the noun's type, and extensive psycholinguistic work (McElree et al. (2001) and Traxler et al. (2002), among others) has demonstrated extra processing costs for metonymic constructions. For example, Traxler et al. (2002) combine metonymic and non-metonymic verbs with entity-denoting and event-denoting nouns (*The boy [started/saw]$_V$ [the puzzle/fight]$_{NP}$*) and report significantly higher processing costs for the "coercion combination" (metonymic verb plus entity-denoting object: *The boy started the puzzle*).

While there has been much debate in theoretical linguistics on individual verbs that may or may not give rise to logical metonymy (for example, on *enjoy*, see Pustejovsky (1995); Fodor and Lepore (1998); Lascarides and Copestake (1998)), work in psycholinguistics (McElree et al., 2001; Traxler et al., 2002; Pylkkänen and McElree, 2006) and computational modeling (Lapata et al., 2003; Lapata and Lascarides, 2003) seem to have agreed on a small set of "metonymic verbs" which is used when looking for empirical correlates of logical metonymy. However, this set of metonymic verbs is semantically rather heterogeneous, as it is selected based on intuition only. It includes not only aspectual verbs[2] (*begin*, *complete*, *continue*, *end*, *finish*, *start*) but also psychological verbs (*enjoy*, *hate*, *like*, *love*, *regret*, *savor*, *try*), as well as others that elude straightforward categorization (*attempt*, *endure*, *manage*, *master*, *prefer*).

---

[1]In this paper we follow the accepted broad linguistic-philosophical distinction between "events" and "(physical) objects" (Casati and Varzi, 2010), using the term "entity" to refer to the ontological class of "object" as opposed to "event". This is to avoid confusion with the grammatical function of "object".

[2]We use the terminology of Levin (1993).

This semantic heterogeneity calls into question a homogeneous notion of metonymic verbs. Indeed, recent work by Katsika et al. (2012) notes that "the hypothesis that eventive inferences must be attributed to the same mechanism of building meaning (coercion + type-shifting) [for all metonymic verbs] is too strong". Their eye-tracking study supports the hypothesis that aspectual verbs trigger coercion and processing cost, while psychological predicates (e.g. *enjoy*) do not. This gives rise a key question: *Are all metonymic verbs alike?*

A second potential methodological risk arises from the fact that experiments need to pair metonymic verbs with a control group of non-metonymic verbs. Verbs that are typically used as non-metonymic include *forget*, *recall*, *remember*, *describe*, *praise*, *prepare*, *shelve*, *see*, and *unpack*. The demarcation of metonymic vs. non-metonymic verbs is rarely motivated explicitly and in some cases even seems rather arbitrary. This raises an evident risk of circularity: the definition of logical metonymy relies on the notion of metonymic verbs, but this class is often characterized only in terms of their triggering metonymic shifts. What is needed is a set of independent and principled criteria to approach what we feel is a second crucial question: *What is a metonymic verb?*

In this paper, we make some progress towards answering these questions by proposing a corpus-based measure of *eventhood* that captures the degree to which verbs expect objects that are events rather than entities. This measure is able to: (a) distinguish between aspectual metonymic verbs, non-aspectual metonymic verbs, and non-metonymic verbs, lending support to Katsika et al.'s (2012) argument; (b) provide empirical evidence for or against the choice of materials in psycholinguistic studies of metonymy; (c) serve as a necessary (although not sufficient) indicator of new verbs that might show metonymic behavior.

Plan of the paper. Section 2 describes the the definition of our eventhood measure ϵ and uses it for data exploration. Section 3 characterizes the data used in two psycholinguistic studies on metonymy. Our results show that our measure can distinguish verb classes, reliably predicting participants' behavior in the experiments.

2   Measuring the Event Expectations of Verbs

Our starting point is that metonymic verbs should be statistically more associated with event-denoting objects, while the non-metonymic verbs should mainly co-occur with entity-denoting objects. We move on to define a measure of "eventhood" of a verb's object slot[3] and to use it to distinguish between verb classes. Our hypotheses are that (a) aspectual verbs have a higher eventhood score than entity-selecting verbs and (b) aspectual verbs have a higher eventhood-score than non-aspectual metonymic verbs.

2.1   Selection of typical objects from corpus data

There has been much work on modeling the various fillers of verbs, i.e. their *selectional preferences*, using explicit or implicit generalizations of the fillers. These rely either primarily on a lexical hierarchy (Resnik, 1996), distributional information (Rooth et al., 1999; Erk et al., 2010) or both (Schulte im Walde et al., 2008). While such computationally-intensive approaches have proven effective in modeling selectional preferences in general, we are interested in learning about only one aspect of a verb's argument, namely how 'event-like' it is.

We use the WordNet (Fellbaum, 2010)[4] lexical hierarchy to discover whether a noun has an event sense. We also use Distributional Memory (DM, Baroni and Lenci (2010)) as a source of distributional information that allows us to determine how strongly a noun is associated with a given verb as an object filler. DM is a general distributional semantic resource which allows the generation of vector-based semantic models (Turney and Pantel, 2010) from the distribution of words in context. In general, distributional semantic models are two-dimensional, relating a word with other words in its context giving

---

[3]We will subsequently simplify the terminology and speak of a "verb's eventhood."

[4]We use version 3 of WordNet, available at: `http://wordnet.princeton.edu/wordnet/download`.

a 'bag-of-words' model (Schütze (1993), cf. Table 1 (a)); or with particular syntactic patterns to give a 'structured vector space' (Padó and Lapata (2007), cf. Table 1 (b)).

$$
\begin{array}{ll}
\text{(a)} \quad \textbf{\textit{dog}}: & \textit{cat}:40.2\ \textit{bone}:25.1\ \textit{best}:10.3\ \ldots \\
\quad\quad\ \ \textbf{\textit{cat}}: & \textit{milk}:37.3\ \ldots \\
\quad\quad\ \ \ \vdots \\
\text{(b)} \quad \textbf{\textit{dog}}: & \langle \textit{obj}, \textit{pet} \rangle:30.2\ \langle \textit{subj}, \textit{bark} \rangle:20.4\ \langle \textit{subj}, \textit{bite} \rangle:7.5\ \ldots \\
\quad\quad\ \ \textbf{\textit{cat}}: & \langle \textit{subj}, \textit{purr} \rangle:25.2\ \ldots \\
\quad\quad\ \ \ \vdots
\end{array}
$$

Table 1: Examples of a two-dimensional bag-of-words space (a), and a two-dimensional structured space (b).

DM is a three-dimensional extension of such a two-dimensional matrix which includes the syntactically derived relation between the two words as an extra dimension. It is derived from the concatenation of the ukWaC[5], the English Wikipedia[6], and the BNC[7], resulting altogether in a 2.83 billion-token corpus. We use the TypeDM variant of DM,[8] which contains over 130M links between nouns, verbs and adjectives, covering generic syntactic relations as well as lexicalized relations (see Baroni and Lenci (2010) for details). In DM, each triple of words w_1, w_2 and relation r, $\langle w_1\ r\ w_2 \rangle$, is scored by the Local Mutual Information (LMI, Evert (2005), Equation 1) between its three elements. LMI contains two factors, (i) the point-wise mutual information which indicates how strongly their co-occurrence deviates from chance and (ii) the raw co-occurrence frequency:

$$
LMI = O_{w_1,r,w_2} \cdot log \frac{O_{w_1,r,w_2}}{E_{w_1,r,w_2}}, \tag{1}
$$

where $E_.$ is the MLE-expected frequency of the triple, and $O_.$ its actually observed frequency in the corpus. For example, since the LMI score for $\langle \textit{meeting obj postpone} \rangle$ is greater than that for $\langle \textit{breakfast obj postpone} \rangle$ we can say that *breakfast* is a less typical object for *postpone* than *meeting*. Defined in this manner, typicality is not only a function of the co-occurrence frequency between an object and a verb but of the significance of this co-occurrence compared to chance. The format of DM allows for the simple extraction of highly informative fillers for a given verb by selecting those tuples whose relation is the one of interest and sorting by score. In the following sections we will use the standard matricization of DM (W×LW) as a semantic space, which defines as dimensions the pairs of links and context words as in Table 1 (b).

2.2 Defining event nouns in a lexical hierarchy

In order to determine how event-like the typical object is for a given predicate, we have to distinguish which objects have an event sense. We define an event noun as a noun with at least one WordNet synset (Fellbaum, 2010) that is dominated in the synset hierarchy by one of the top nodes shown in Table 2. This is a simple approximation of the degree to which the noun denotes an event. A more informed measure could e.g. include distributional information of the noun's senses. It is important to note that a particular noun can have more than one event-dominated synset. There are in fact eight nouns whose synsets generalize to all of the event nodes designated:

control, culture, differentiation, elimination, inspiration, pleasure, reproduction, rumination,

that is, they all have an *action, cognitive process*, and *biological process* reading.

---

[5] http://wacky.sslmit.unibo.it

[6] http://en.wikipedia.org

[7] http://www.natcorp.ox.ac.uk

[8] TypeDM is available from http://clic.cimec.unitn.it/dm/.

WordNet node	Count	Examples
EVENT	11248	*training, splat, Alamo, suicide, hyperalimentation*
ACT/DEED/HUMAN ACTION/HUMAN ACTIVITY, ACTION, ACTIVITY	9845	*banditry, dissolution, beanball, messaging, banishment*
PROCESS/PHYSICAL PROCESS	2590	*ultracentrifugation, desensitisation, extinction, superconductivity*
PROCESS/COGNITIVE PROCESS/MENTAL PROCESS /OPERATION/COGNITIVE OPERATION	998	*reminiscence, breakdown, score, analogy, inference*
ORGANIC PROCESS/BIOLOGICAL PROCESS	878	*recuperation, emission, autoregulation, drinking, blossoming*
all	14143	

Table 2: High-level event-denoting nodes in WordNet with examples.

This definition leads to a set EV of 14K event nouns (out of WordNet's 170K nouns), which we can use to determine to what extent 'the typical object' of a verb is event-like. First we take the k most strongly associated object fillers from DM, $obj_k(v)$ for the verb v and then define the eventhood to be the percentage of these fillers that have an event sense. In other words, the eventhood ϵ_k for a verb v is defined as:

$$\epsilon_k(v) = \frac{|EV \cap obj_k(v)|}{k}. \tag{2}$$

Selecting the top k scored fillers as prototypical arguments has proven a reliable method to characterize the expectations for the argument slot which allows, e.g., the modeling of selectional preferences (cf. Baroni and Lenci (2010); Erk et al. (2010); Lenci (2011)). For the present analysis, we fix k at 100 (i.e. $\epsilon := \epsilon_{100}$), we thereby also eliminate the issue of using words from DM which are not covered in WordNet. The following section investigates the range of eventhood scores across the verbs in DM.

2.3 Evaluation on Verbs in DM

Figure 1 shows the distribution of eventhood across verbs in DM. Verbs with $\epsilon \approx 0$, i.e. verbs with low eventhood, include *unfrock, detain, marry,* and *behead*, while verbs with high eventhood, i.e. those which rank the highest with respect to ϵ (i.e. $\epsilon \approx 1$), include *expedite, undergo, halt,* and *delay*.
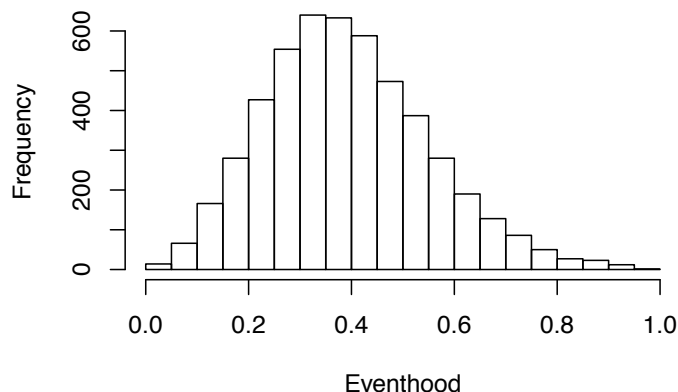

Figure 1: Histogram of eventhood across verbs in DM.

While this 'linearization' of the space of verbs given by their eventhood scores does not in and of itself suggest semantic coherence—given a particular range $[\alpha; \beta]$, the class of verbs with $\alpha < \epsilon(v) < \beta$ will in general be a heterogenous class—we find that in the fringe ranges, i.e. where $\epsilon \approx 0$ and $\epsilon \approx 1$, the verbs

appear to be coherent with respect to their object fillers. For instance, the left most bar in the histogram corresponding to the range $0 < \epsilon < 0.5$ typically have people as the experiencers of the action denoted by the verb. In a sense, things that happen to or with people (e.g. *marry* or *behead*) do not typically happen to or with events. On the other side of the spectrum we have only 13 verbs with $0.9 < \epsilon < 1$ (e.g. *commence, cease, halt, delay*), most of which concern the temporal unfolding of an event.
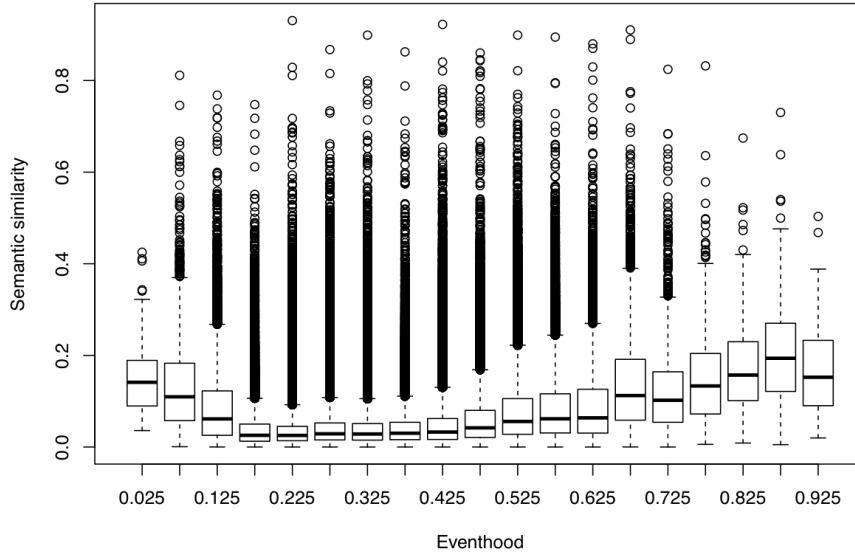

Figure 2: Pairwise semantic similarity within $\{v | \alpha < \epsilon(v) < \beta\}$ in DM.

The most frequent range ($0.3 < \epsilon < 0.35$), covering 640 verbs, contains a very diverse group of verbs: *prance, fluoridate, emaciate* ($\epsilon \approx 0.3$) to *exorcise, downsize, muddy* ($\epsilon \approx 0.35$). To determine the semantic coherence across eventhood scores, we computed the pairwise cosine semantic similarity between the verbs within each eventhood range (Figure 1). Figure 2 shows the semantic similarities among verbs for each bin in the range $[\alpha; \beta]$. The similarities for each set of n verbs $\{v | \alpha < \epsilon(v) < \beta\}$ were then contrasted with the pairwise similarities for n randomly drawn verbs. In 19 out of the 20 bins the actual verb similarities were statistically higher than the random ones ($p < .001$). This means that the verbs within each range form a semantically coherent group, suggesting that the eventhood score can identify semantically related verbs.

Towards either end of the eventhood spectrum (Figure 2), we see that the verbs are semantically much more similar to one another, while the mid-range is the most semantically dissimilar. In the extreme cases, we are dealing with verbs that are similar to one another, while in the mid-range the semantic coherence is lost.

3 Evaluation on Psycholinguistic Datasets

We test our model on the experimental datasets from two metonymy interpretation studies (Traxler et al., 2002; Katsika et al., 2012). Each of these studies makes use of a classification, according to which it expects participants' behavior to differ. More precisely, they expect more difficulty in processing when 'metonymic' verbs are combined with non-event denoting objects than when 'less metonymic' verbs are.

If, as those studies claim, event-selecting verbs give rise to higher processing costs when combined with entity-denoting objects, then we expect our eventhood measure to be able to distinguish between the classes used in the psycholinguistic studies.

Traxler et al.'s (2002) dataset	metonymic	begin, complete, end, endure, enjoy, expect, finish, prefer, start
	non-metonymic	approve, curse, describe, forget, ignore, observe, outline, praise, prepare, recall, recollect, remember, report, see, watch
Katsika et al.'s (2012) dataset	metonymic aspectual	begin, complete, continue, finish, start
	metonymic psychological	enjoy, face, favor, prefer, resist, stomach, tolerate
	non-metonymic	access, auction, buy, conduct, contribute, deliver, destroy, drop, fax, find, inspect, misplace, open, peruse, purchase, photocopy, rent, sell, send, shelve, shred, submit, trash, unearth, unpack, write

Table 3: Datasets from Traxler et al. (2002) and Katsika et al. (2012).

3.1 The Datasets

The two datasets used are:

Traxler et al. (2002) dataset: composed of 24 verbs used in Experiment 2 and 3 in Traxler et al. (2002). Verbs are divided in *metonymic* and *non-metonymic* verbs (*event verbs* and *neutral verbs*, according to the terminology of the study). Higher processing costs were yielded for metonymic verbs combined with entity-denoting objects than for all remaining conditions (metonymic verbs combined with event-denoting objects and non-metonymic verbs combined with entity and event-denoting objects).

Katsika et al. (2012) dataset: composed of 38 verbs used in Katsika et al. (2012) taken mostly from previous psycholinguistic experiments on type-shifting.[9] As mentioned above, Katsika et al. (2012) make a point of distinguishing between three sets of verbs: here *metonymic aspectual*, *metonymic psychological* and *non-metonymic* verbs. (according to the terminology of the study, *aspectual*, *psychological* and *entity-selecting*). Readers spent more time re-reading the verb in the metonymic aspectual condition than the metonymic psychological or non-metonymic condition.

3.2 Evaluation

A direct correlation between eventhood and reading times is not feasible, because the psycholinguistic studies do not report reading times for each verb, but rather the average per condition (and even if they did, the number of measurements per verb would be too small). Thus, we resort to two alternative evaluation methods:

1. For both datasets, we report the Wilcoxon rank sum test (a non-parametric analog of Student's t-test) to check for differences in eventhood between verb classes.

2. In Traxler et al.'s (2002) dataset, each sentence exists once with a metonymic verb and once with a non-metonymic verb, which gives us a list of verb pairs. This list allows us to compute the number of times the eventhood of the metonymic verb is higher than the eventhood of the non-metonymic verb.

---

[9]The study also used verbs that do not select for a direct object. We excluded these.

| metonymic | | non-metonymic | | prediction |
verb	eventhood	verb	eventhood	correct?
begin	0.91	praise	0.55	y
complete	0.79	recall	0.67	y
start	0.78	see	0.51	y
endure	0.73	report	0.78	n
end	0.72	outline	0.64	y
finish	0.66	prepare	0.41	y
enjoy	0.57	watch	0.60	n
enjoy	0.57	curse	0.31	y
prefer	0.54	praise	0.55	n

Table 4: Eventhood values for some verb pairs from Traxler et al. (2002) and correctness of model prediction.

3.3 Results

We first consider the Katsika et al. (2012) data. Metonymic aspectual verbs yield higher eventhood scores compared to metonymic psychological verbs and non-metonymic verbs. All pairwise comparisons are significant: metonymic aspectual verbs vs. metonymic psychological verbs ($W = 30, p < 0.05$); metonymic aspectual verbs vs. non-metonymic verbs ($W = 125, p < 0.01$); metonymic psychological vs. non-metonymic verbs ($W = 18.5, p < 0.01$).

For the Traxler et al. (2002) dataset, the difference between metonymic verbs and non-metonymic verbs is close to significance, with p just above 0.05 ($W = 100.5, p < 0.053$). The fact that this difference is less significant is compatible with the observations in Katsika et al. (2012), namely that the set of verbs typically used in studies on logical metonymy is heterogeneous and includes verbs which are less event-selecting than aspectual verbs. In fact, if we remove the four metonymic verbs that are not aspectual (*endure, enjoy, expect, prefer*), we find a significant difference between the non-metonymical and metonymic (now aspectual-only) classes ($W = 67.5, p < 0.01$).

On the Traxler et al. (2002) dataset, the model scores 23/32 in the pairwise comparisons. In other words, metonymic verbs receive higher eventhood scores for 72 % of the pairs. Table 4 shows some examples of the pairwise comparisons. We find that errors tend to occur for metonymic psychological verbs more often than for metonymic aspectual verbs. The reason is that the most event-affine non-metonymic verbs (*recall, report*) prefer events to a higher degree than the least event-affine metonymic verbs (*enjoy, prefer*). This again suggests that Traxler et al.'s (2002) set of metonymic verbs is not clearly distinct from their non-metonymic verbs. This point is reinforced by Figure 3 which visualizes the eventhood distributions for the verb classes in both datasets as density plots and boxplots. The more homogeneous three-class distinctions in Katsika et al. (2012) seems justified as it clearly identifies three different selection behaviors (metonymic aspectual, metonymic psychological, non-metonymic), while the two-class distinction in Traxler et al. (2002) shows substantial overlap.

3.4 Discussion

Our results indicate that eventhood is a good indicator of 'metonymicity' and can even distinguish between classes of metonymic verbs. This raises the question of how strong the correlation between metonymicity and eventhood really is.

A first question is whether verbs need to have an (almost) perfect eventhood score to be metonymic. This is not plausible: if a verb is metonymic, we expect it to allow for entity-denoting objects, even if they will occur less frequently. For instance, *begin* is, arguably, a 'true' metonymic verb (metonymic aspectual). However, occurrences of *begin* in metonymical context (with entity objects) are indeed attested in the
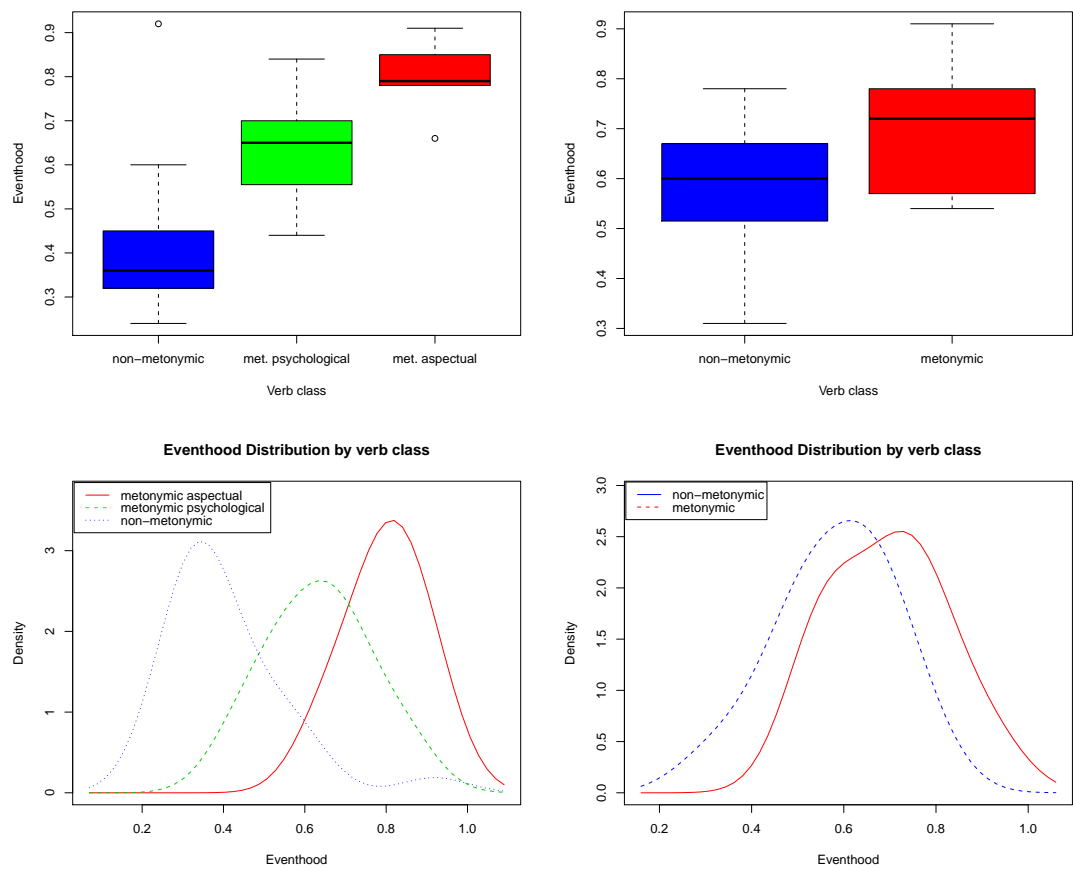
Figure 3: Comparing eventhood distribution for verb classes in the Katsika et al. (2012) dataset (left) and Traxler et al. (2002) dataset (right).

corpus. Consequently, it obtains an eventhood score of 0.91. Generally, we expect metonymic verbs to be placed at the high end of the eventhood spectrum, but not at the extreme (cf. Figure 3).

A second question is whether all verbs at the upper end of the eventhood range are (or at least can be) metonymic. Our inspection shows that verbs with an extremely high eventhood tend to disprefer metonymic constructions. Among the top eventhood-scoring verbs are, for instance, *perform, undergo, protest, conduct, spearhead, facilitate, undertake, witness*. All of these verbs clearly prefer events and occur infrequently in metonymic constructions. However, occasional metonymic productivity occurs, as in the following examples from American discussion forums on the web:

There's a huge connection between Prematurity and GBS morbidities and mortalities and I too would be more then willing to underfine the antibiotics if such a risk factor was involved.

[The Adventures of Tom Sawyer] is called the first real work of the American Literature movement, which in general spawned the Hemingways and Faulkners I would later undertake.

Taking an IPD approach, we collaborated with Zeemac using 3D modeling known as "real time design" to facilitate the floor plan.

In sum, the correlation between eventhood and metonymicity is strong but not perfect. It remains a question for further investigation which other factors are involved in determining whether a high-eventhood verb features prominently in metonymic constructions (*begin*) or not (*conduct*). One factor that we want to

investigate is *specificity*, following the intuition that only verbs that refer to general properties of as many events as possible (like aspectual properties) rather than specific scenarios are suitable as metonymic verbs.

4 Conclusions

In this paper, we have introduced a simple data-driven measure of *eventhood*, that is, the degree to which verbs prefer events over entities as their direct objects. Our eventhood measure allows us to characterize and separate verb classes relevant for logical metonymy that were so far hand-picked on the basis of intuitive considerations.

The fundamentally graded nature of our measure suggests that there is no clear-cut binary distinction between metonymic verbs on one end and non-metonymic verbs on the other. Instead, there is a continuum with a sequence of classes (named in decreasing order of eventhood): First, verbs with an extremely high eventhood such as *undergo* strongly disprefer entity-denoting objects, but in some creative uses they may still combine with them giving rise to metonymic interpretation. Next, metonymic aspectual verbs strongly prefer event-denoting objects but are (albeit less frequently) attested with entity-denoting objects and form "classic" cases of metonymy. Psychological verbs have a less strong bias for event-denoting objects, but can still be considered as metonymic (although, as Katsika et al. (2012) argue, with their own behavioral patterns). Finally, there is the wide range of non-metonymic verbs that are either neutral or entity-prefering.

This picture indicates that the question of how to select verbs for the control condition against which metonymical verbs are compared is by no means trivial. We believe that our depiction of the metonymic behavior as a graded range suggests that eventhood can be used to inform and guide the design of further materials in this area.

In closing, we note that expectations for the semantic types (event/entity) of verbal arguments can be understood as a very coarse variant of selectional preferences, and our model as a much simplified version of ontological models of selectional preferences (Resnik, 1996). On the other hand, the existence of classes with graded preferences indicates that eventhood differences may not be binary distinctions, but that we might rather be dealing with a graded range of behaviors. This has clear consequences for type-clash accounts of logical metonymy: given the existence of many verbs which exhibit intermediate behavior, it seems unlikely that there are two exclusive classes (metonymic vs. non-metonymic). Within this graded picture, the function of the type clash may be taken over by mismatches between preference (expectation) for an object and the actually encountered object.

The preliminary investigations presented in this paper thus show that corpus data can be used to provide independent empirical grounding to theory-loaded notions such as the one of metonymic verbs. This can be extremely useful for future experimental work as well as to evaluate experimental results.

References

Baroni, M. and A. Lenci (2010). Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics 36*(4), 1–49.

Casati, R. and A. C. Varzi (2010). Events. In N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/spr2010/entries/events/.

Erk, K., S. Padó, and U. Padó (2010). A flexible, corpus-driven model of regular and inverse selectional preferences. *Computational Linguistics 36*(4), 723–763.

Evert, S. (2005). *The Statistics of Word Co-Occurrences: Word Pairs and Collocations*. Ph. D. thesis, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.

Fellbaum, C. (2010). Wordnet. *Theory and Applications of Ontology: Computer Applications*, 231–243.

Fodor, J. and E. Lepore (1998). The emptiness of the lexicon: reflections on James Pustejovsky's The Generative Lexicon. *Linguistic Inquiry 29*(2), 269–288.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. MIT Press.

Katsika, A., D. Braze, A. Deo, and M. Piñango (2012). Complement coercion: Distinguishing between type-shifting and pragmatic inferencing. *The Mental Lexicon 7*(1), 58–76.

Lapata, M., F. Keller, and C. Scheepers (2003). Intra-sentential context effects on the interpretation of logical metonymy. *Cognitive Science 27*(4), 649–668.

Lapata, M. and A. Lascarides (2003). A probabilistic account of logical metonymy. *Computational Linguistics 29*(2), 263–317.

Lascarides, A. and A. Copestake (1998). Pragmatics and word meaning. *Journal of Linguistics 34*(02), 387–414.

Lenci, A. (2011). Composing and updating verb argument expectations: A distributional semantic model. In *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics*, Portland, Oregon, pp. 58–66.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*, Volume 348. University of Chicago press Chicago, IL.

McElree, B., M. Traxler, M. Pickering, R. Seely, and R. Jackendoff (2001). Reading time evidence for enriched composition. *Cognition 78*(1), B17–B25.

Padó, S. and M. Lapata (2007). Dependency-based construction of semantic space models. *Computational Linguistics 33*(2), 161–199.

Pustejovsky, J. (1995). *The Generative Lexicon*. MIT Press.

Pylkkänen, L. and B. McElree (2006). The syntax-semantic interface: On-line composition of sentence meaning. In *Handbook of Psycholinguistics*, pp. 537–577. Elsevier.

Resnik, P. (1996). Selectional constraints: an information-theoretic model and its computational realization.

Rooth, M., S. Riezler, D. Prescher, G. Carroll, and F. Beil (1999). Inducing a Semantically Annotated Lexicon via EM-Based Clustering. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, Maryland, MD.

Schulte im Walde, S., C. Hying, C. Scheible, and H. Schmid (2008). Combining EM training and the MDL principle for an automatic verb classification incorporating selectional preferences. In *Proc. of ACL*, pp. 496–504.

Schütze, H. (1993). Word space. In *Advances in Neural Information Processing Systems 5*. Citeseer.

Traxler, M. J., M. J. Pickering, and B. McElree (2002). Coercion in sentence processing: evidence from eye-movements and self-paced reading. *Journal of Memory and Language 47*, 530–547.

Turney, P. D. and P. Pantel (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research 37*, 141–188.