

Sense-based Interpretation of Logical Metonymy Using a Statistical Method

Ekaterina Shutova

Computer Laboratory
University of Cambridge
15 JJ Thomson Avenue
Cambridge CB3 0FD, UK

Ekaterina.Shutova@cl.cam.ac.uk

Abstract

The use of figurative language is ubiquitous in natural language texts and it is a serious bottleneck in automatic text understanding. We address the problem of interpretation of *logical metonymy*, using a statistical method. Our approach originates from that of Lapata and Lascarides (2003), which generates a list of non-disambiguated interpretations with their likelihood derived from a corpus. We propose a novel sense-based representation of the interpretation of logical metonymy and a more thorough evaluation method than that of Lapata and Lascarides (2003). By carrying out a human experiment we prove that such a representation is intuitive to human subjects. We derive a ranking scheme for verb senses using an unannotated corpus, WordNet sense numbering and glosses. We also provide an account of the requirements that different aspectual verbs impose onto the interpretation of logical metonymy. We tested our system on verb-object metonymic phrases. It identifies and ranks metonymic interpretations with the mean average precision of 0.83 as compared to the gold standard.

1 Introduction

Metonymy is defined as the use of a word or a phrase to stand for a related concept which is not explicitly mentioned. Here are some examples of metonymic phrases:

- (1) The *pen* is mightier than the *sword*.
- (2) He played *Bach*.

- (3) He drank *his glass*. (Fass, 1991)
- (4) He enjoyed *the book*. (Lapata and Lascarides, 2003)
- (5) After *three martinis* John was feeling well. (Godard and Jayez, 1993)

The metonymic adage in (1) is a classical example. Here the *pen* stands for the press and the *sword* for military power. In the following example *Bach* is used to refer to the composer's music and in (3) the *glass* stands for its *content*, i.e. the actual *drink* (beverage).

The sentences (4) and (5) represent a variation of this phenomenon called *logical metonymy*. Here both *the book* and *three martinis* have eventive interpretations, i.e. the noun phrases stand for the events of *reading the book* and *drinking three martinis* respectively. Such behaviour is triggered by the type requirements the verb (or the preposition) places onto its argument. This is known in linguistics as a phenomenon of *type coercion*. Many existing approaches to logical metonymy explain systematic syntactic ambiguity of metonymic verbs (such as *enjoy*) or prepositions (such as *after*) by means of type coercion (Pustejovsky, 1991; Pustejovsky, 1995; Briscoe et al., 1990; Verspoor, 1997; Godard and Jayez, 1993).

Logical metonymy occurs in natural language texts relatively frequently. Therefore, its automatic interpretation would significantly facilitate the task of many NLP applications that require semantic processing (e.g., machine translation, information extraction, question answering and many others). Utiyama et al. (2000) followed by Lapata and Lascarides (2003) used text corpora to automatically derive interpretations of metonymic phrases.

Utiyama et al. (2000) used a statistical model for the interpretation of general metonymies for Japanese. Given a verb-object metonymic phrase, such as *read Shakespeare*, they searched for entities the object could stand for, such as *plays of Shakespeare*. They considered all the nouns co-occurring with the object noun and the Japanese equivalent of the preposition *of*. Utiyama and his colleagues tested their approach on 75 metonymic phrases taken from the literature and reported a precision of 70.6%, whereby an interpretation was considered correct if it made sense in some imaginary context.

Lapata and Lascarides (2003) extend Utiyama’s approach to interpretation of logical metonymies containing aspectual verbs (e.g. *begin the book*) and polysemous adjectives (e.g. *good meal* vs. *good cook*). Their method generates a list of interpretations with their likelihood derived from a corpus.

Lapata and Lascarides define an interpretation of logical metonymy as a verb string, which is ambiguous with respect to word sense. Some of these strings indeed correspond to paraphrases that a human would give for the metonymic phrase. But they are not meaningful as such for automatic processing, since their senses still need to be disambiguated in order to obtain the actual meaning. For example, compare the *grab* sense of *take* vs. its *film* sense for the metonymic phrase *finish video*. It is obvious that only the latter sense is a correct interpretation.

We extend the experiment of Lapata and Lascarides by disambiguating the interpretations with respect to WordNet (Fellbaum, 1998) synsets (for verb-object metonymic phrases). We propose a novel ranking scheme for the synsets using a non-disambiguated corpus, address the issue of sense frequency distribution and utilize information from WordNet glosses to refine the ranking.

We conduct an experiment to show that our representation of a metonymic interpretation as a synset is intuitive to human subjects. In the discussion section we provide an overview of the constraints on logical metonymy pointed out in linguistics literature, as well as proposing some additional constraints (e.g. on the type of the metonymic verb, on the type of the reconstructed event, etc.)

| Metonymic Phrase | Interpretations | Log-probability |
|------------------|-----------------|-----------------|
| finish video | film | -19.65 |
| | edit | -20.37 |
| | shoot | -20.40 |
| | view | -21.19 |
| | play | -21.29 |
| | stack | -21.75 |
| | make | -21.95 |
| | programme | -22.08 |
| | pack | -22.12 |
| | use | -22.23 |
| | watch | -22.36 |
| | produce | -22.37 |

Table 1: Interpretations of Lapata and Lascarides (2003) for *finish video*

2 Lapata and Lascarides’ Method

The intuition behind the approach of Lapata and Lascarides is similar to that of Pustejovsky (1991; 1995), namely that there is an event not explicitly mentioned, but implied by the metonymic phrase (*begin to read the book*, or *the meal that tastes good* vs. *the cook that cooks well*). They used the British National Corpus (BNC)(Burnard, 2007) parsed by the Cass parser (Abney, 1996) to extract events (verbs) co-occurring with both the metonymic verb (or adjective) and the noun independently and ranked them in terms of their likelihood according to the data. The likelihood of a particular interpretation is calculated using the following formula:

$$P(e, v, o) = \frac{f(v, e) \cdot f(o, e)}{N \cdot f(e)}, \quad (1)$$

where e stands for the eventive interpretation of the metonymic phrase, v for the metonymic verb and o for its noun complement. $f(e)$, $f(v, e)$ and $f(o, e)$ are the respective corpus frequencies. $N = \sum_i f(e_i)$ is the total number of verbs in the corpus. The list of interpretations Lapata and Lascarides (2003) report for the phrase *finish video* is shown in Table 1.

Lapata and Lascarides compiled their test set by selecting 12 verbs that allow logical metonymy¹ from the lexical semantics literature and combining each of them with 5 nouns. This yields 60 phrases, which were then manually filtered, excluding 2 phrases as non-metonymic.

They compared their results to paraphrase judgements elicited from humans. The subjects were presented with three interpretations for each

¹*attempt, begin, enjoy, finish, expect, postpone, prefer, resist, start, survive, try, want*

metonymic phrase (from high, medium and low probability ranges) and were asked to associate a number with each of them reflecting how good they found the interpretation. They report a correlation of 0.64, whereby the inter-subject agreement was 0.74. It should be noted, however, that such an evaluation scheme is not very informative as Lapata and Lascarides calculate correlation only on 3 data points for each phrase out of many more yielded by the model. It fails to take into account the quality of the list of top interpretations, although the latter is deemed to be the aim of such applications. In comparison the fact that Lapata and Lascarides initially select the interpretations from high, medium or low probability ranges makes the task significantly easier.

3 Alternative Interpretation of Logical Metonymy

The approach of Lapata and Lascarides (2003) produces a list of non-disambiguated verbs, essentially just strings, representing possible interpretations of a metonymic phrase. We propose an alternative representation of metonymy interpretation consisting of a list of senses that map to WordNet synsets. However, the sense-based representation builds on the list of non-disambiguated interpretations similar to the one of Lapata and Lascarides.

Our method consists of the following steps:

- **Step 1** Use the method of Lapata and Lascarides (2003) to obtain a set of candidate interpretations (strings) from a non-annotated corpus. We expect our reimplementing of the method to extract data more accurately, since we use a more robust parser (RASP (Briscoe et al., 2006)), take into account more syntactic structures (coordination, passive), and extract our data from a newer version of the BNC.
- **Step 2** Map strings to WordNet synsets. We noticed that good interpretations in the lists yielded by Step 1 tend to form coherent semantic classes (e.g. *take, shoot [a video]* vs. *view, watch [a video]*). We search the list for verbs, whose senses are in hyponymy and synonymy relations with each other according to WordNet and store these senses.
- **Step 3** Rank the senses, adopting Zipfian sense frequency distribution and using the

initial string likelihood as well as the information from WordNet glosses.

Sense disambiguation is essentially performed in both Step 2 and Step 3. One of the challenges of our task is that we use a non-disambiguated corpus while ranking particular senses. This is due to the fact that there is no word sense disambiguated corpus available, which would be large enough to reliably extract statistics for metonymic interpretations.

4 Extracting Ambiguous Interpretations

4.1 Parameter Estimation

We used the method developed by Lapata and Lascarides (2003) to create the initial list of non-disambiguated interpretations. The parameters of the model were estimated from the British National Corpus (BNC) (Burnard, 2007) that was parsed using the RASP parser of Briscoe et al. (2006). We used the grammatical relations (GRs) output of RASP for BNC created by Andersen et al. (2008). In particular, we extracted all direct and indirect object relations for the nouns from the metonymic phrases, i.e. all the verbs that take the head noun in the compliment as an object (direct or indirect), in order to obtain the counts for $f(o, e)$. Relations expressed in the passive voice and with the use of coordination were also extracted. The verb-object pairs attested in the corpus only once were discarded, as well as the verb *be*, since it does not add any semantic information to the metonymic interpretation. In the case of indirect object relations, the verb was considered to constitute an interpretation together with the preposition, e.g. for the metonymic phrase *enjoy the city* the correct interpretation is *live in* as opposed to *live*.

As the next step we need to identify all possible verb phrase (VP) complements to the metonymic verb (both progressive and infinitive), which represent $f(v, e)$. This was done by searching for `xcomp` relations in the GRs output of RASP, in which our metonymic verb participates in any of its inflected forms. Infinitival and progressive complement counts were summed up to obtain the final frequency $f(v, e)$.

After the frequencies $f(v, e)$ and $f(o, e)$ were obtained, possible interpretations were ranked according to the model of Lapata and Lascarides (2003). The top interpretations for the metonymic

| finish video | | enjoy book | |
|-----------------|----------|-----------------|----------|
| Interpretations | Log-prob | Interpretations | Log-prob |
| view | -19.68 | read | -15.68 |
| watch | -19.84 | write | -17.47 |
| shoot | -20.58 | work on | -18.58 |
| edit | -20.60 | look at | -19.09 |
| film on | -20.69 | read in | -19.10 |
| film | -20.87 | write in | -19.73 |
| view on | -20.93 | browse | -19.74 |
| make | -21.26 | get | -19.90 |
| edit of | -21.29 | re-read | -19.97 |
| play | -21.31 | talk about | -20.02 |
| direct | -21.72 | see | -20.03 |
| sort | -21.73 | publish | -20.06 |
| look at | -22.23 | read through | -20.10 |
| record on | -22.38 | recount in | -20.13 |

Table 2: Possible Interpretations of Metonymies Ranked by our System

phrases *enjoy book* and *finish video* together with their log-probabilities are shown in Table 2.

4.2 Comparison with the Results of Lapata and Lascarides

We compared the output of our reimplementation of Lapata and Lascarides’ algorithm with their results, which we obtained from the authors. The major difference between the two systems is that we extracted our data from the BNC parsed by RASP, as opposed to the Cass chunk parser (Abney, 1996) utilized by Lapata and Lascarides. Our system finds approximately twice as many interpretations as theirs and covers 80% of their lists (our system does not find some of the low-probability range verbs of Lapata and Lascarides). We compared the rankings of the two implementations in terms of Pearson correlation coefficient and obtained the average correlation of 0.83 (over all metonymic phrases).

We also evaluated the performance of our system against the judgements elicited from humans in the framework of the experiment of Lapata and Lascarides (2003) (for a detailed description of the human evaluation setup see (Lapata and Lascarides, 2003), pages 12-18). The Pearson correlation coefficient between the ranking of our system and the human ranking equals to 0.62 (the intersubject agreement on this task is 0.74). This is slightly lower than the number achieved by Lapata and Lascarides (0.64). Such a difference is probably due to the fact that our system does not find some of the low-probability range verbs that Lapata and Lascarides included in their test set, and thus those interpretations get assigned a probability of 0. We conducted a one-tailed t-test to

determine if our counts were significantly different from those of Lapata and Lascarides. The difference is statistically insignificant ($t=3.6$; $df=180$; $p<.0005$), and the output of the system is deemed acceptable to be used for further experiments.

5 Mapping Interpretations to WordNet Senses

The interpretations at this stage are just strings representing collectively all senses of the verb. What we aim for is the list of verb senses that are correct interpretations for the metonymic phrase. We assume the WordNet synset representation of a sense.

It has been recognized (Pustejovsky, 1991; Pustejovsky, 1995; Godard and Jayez, 1993) and verified by us empirically that correct interpretations tend to form semantic classes, and therefore, correct interpretations should be related to each other by semantic relations, such as synonymy or hyponymy. In order to select the right senses of the verbs in the context of the metonymic phrase we did the following.

- We searched the WordNet database for the senses of the verbs that are in synonymy, hyponymy and hyponym relations.
- We stored the corresponding synsets in a new list of interpretations. If one synset was a hyponym (or hyponym) of the other, then both synsets were stored.

For example, for the metonymic phrase *finish video* the interpretations *watch*, *view* and *see* are synonymous, therefore a synset containing (*watch(3) view(3) see(7)*) was stored. This means that sense 3 of *watch*, sense 3 of *view* and sense 7 of *see* would be correct interpretations of the metonymic expression.

The obtained number of synsets ranges from 14 (*try shampoo*) to 1216 (*want money*) for the whole dataset of Lapata and Lascarides (2003).

6 Ranking the Senses

A problem that arises with the lists of synsets obtained is that they contain different senses of the same verb. However, very few verbs have such a range of meanings that their two different senses could represent two distinct metonymic interpretations (e.g., in case of *take* interpretation of *finish video shoot* sense and *look at, consider* sense are

both acceptable interpretations, the second obviously being dispreferred). In the vast majority of cases the occurrence of the same verb in different synsets means that the list still needs filtering.

In order to do this we rank the synsets according to their likelihood of being a metonymic interpretation. The sense ranking is largely based on the probabilities of the verb strings derived by the model of Lapata and Lascarides (2003).

6.1 Zipfian Sense Frequency Distribution

The probability of each string from our initial list represents the sum of probabilities of all senses of this verb. Hence this probability mass needs to be distributed over senses first. The sense frequency distribution for most words tends to be closer to Zipfian, rather than uniform or any other distribution (Preiss, 2006). This is an approximation that we rely on, as it has been shown to realistically describe the majority of words.

This means that the first senses will be favoured over the others, and the frequency of each sense will be inversely proportional to its rank in the list of senses (i.e. sense number, since word senses are ordered in WordNet by frequency).

$$P_{v,k} = P_v \cdot \frac{1}{k} \quad (2)$$

where k is the sense number and P_v is the likelihood of the verb string being an interpretation according to the corpus data, i.e.

$$P_v = \sum_{k=1}^{N_v} P_{v,k} \quad (3)$$

where N_v is the total number of senses for the verb in question.

The problem that arises with (2) is that the inverse sense numbers ($1/k$) do not add up to 1. In order to circumvent this, the Zipfian distribution is commonly normalised by the N th generalised harmonic number. Assuming the same notation

$$P_{v,k} = P_v \cdot \frac{1/k}{\sum_{n=1}^{N_v} 1/n} \quad (4)$$

Once we have obtained the sense probabilities $P_{v,k}$, we can calculate the likelihood of the whole synset

$$P_s = \sum_{i=1}^{I_s} P_{v_i,k} \quad (5)$$

where v_i is a verb in the synset s and I_s is the total number of verbs in the synset s . The verbs suggested by WordNet, but not attested in the corpus in the required environment, are assigned the probability of 0. Some output synsets for the metonymic phrase *finish video* and their log-probabilities are demonstrated in Table 3.

In our experiment we compare the performance of the system assuming a Zipfian distribution of senses against the baseline using a uniform distribution. We expect the former to yield better results.

6.2 Gloss Processing

The model in the previous section penalizes synsets that are incorrect interpretations. However, it can not discriminate well between the ones consisting of a single verb. By default it favours the sense with a smaller sense number in WordNet. This poses a problem for the examples such as *direct* for the phrase *finish video*: our list contains several senses of it, as shown in Table 4, and their ranking is not satisfactory. The only correct interpretation in this case, sense 3, is assigned a lower likelihood than the senses 1 and 2.

The most relevant synset can be found by using the information from WordNet glosses (the verbal descriptions of concepts, often with examples). We searched for the glosses containing terms related to the noun in the metonymic phrase, here *video*. Such related terms would be its direct synonyms, hyponyms, hypernyms, meronyms or holonyms according to WordNet. We assigned more weight to the synsets whose gloss contained related terms. In our example the synset (*direct-v-3*), which is the correct metonymic interpretation, contained the term *film* in its gloss and was therefore selected. Its likelihood was multiplied by the factor of 10.

It should be noted, however, that the glosses do not always contain the related terms; the expectation is that they will be useful in the majority of cases, not in all of them.

7 Evaluation

7.1 The Gold Standard

We selected the most frequent metonymic verbs for our experiments: *begin*, *enjoy*, *finish*, *try*, *start*. We randomly selected 10 metonymic phrases containing these verbs. We split them into the development set (5 phrases) and the test set (5 phrases)

| Synset and its Gloss | Log-prob |
|--|----------|
| (watch-v-1) - look attentively; "watch a basketball game" | -4.56 |
| (view-v-2 consider-v-8 look-at-v-2) - look at carefully; study mentally; "view a problem" | -4.66 |
| (watch-v-3 view-v-3 see-v-7 catch-v-15 take-in-v-6) - see or watch; "view a show on television"; "This program will be seen all over the world"; "view an exhibition"; "Catch a show on Broadway"; "see a movie" | -4.68 |
| (film-v-1 shoot-v-4 take-v-16) - make a film or photograph of something; "take a scene"; "shoot a movie" | -4.91 |
| (edit-v-1 redact-v-2) - prepare for publication or presentation by correcting, revising, or adapting; "Edit a book on lexical semantics"; "she edited the letters of the politician so as to omit the most personal passages" | -5.11 |
| (film-v-2) - record in film; "The coronation was filmed" | -5.74 |
| (screen-v-3 screen-out-v-1 sieve-v-1 sort-v-1) - examine in order to test suitability; "screen these samples"; "screen the job applicants" | -5.91 |
| (edit-v-3 cut-v-10 edit-out-v-1) - cut and assemble the components of; "edit film"; "cut recording tape" | -6.20 |

Table 3: Metonymy Interpretations as Synsets (for *finish video*)

| Synset and its Gloss | Log-prob |
|---|----------|
| (direct-v-1) - command with authority; "He directed the children to do their homework" | -6.65 |
| (target-v-1 aim-v-5 place-v-7 direct-v-2 point-v-11) - intend (something) to move towards a certain goal; "He aimed his fists towards his opponent's face"; "criticism directed at her superior"; "direct your anger towards others, not towards yourself" | -7.35 |
| (direct-v-3) - guide the actors in (plays and films) | -7.75 |
| (direct-v-4) - be in charge of | -8.04 |

Table 4: Different Senses of *direct* (for *finish video*)

| Development Set | Test Set |
|------------------|----------------|
| enjoy book | enjoy story |
| finish video | finish project |
| start experiment | try vegetable |
| finish novel | begin theory |
| enjoy concert | start letter |

Table 5: Metonymic Phrases in Development and Test Sets

given in the table 5.

The gold standards were created for the top 30 synsets of each metonymic phrase after ranking. This threshold was set experimentally: the recall of correct interpretations among the top 30 synsets is 0.75 (average over metonymic phrases from the development set). This threshold allows to filter out a large number of incorrect interpretations.

The interpretations that are plausible in some imaginary context are marked as correct in the gold standard.

7.2 Evaluation Measure

We evaluated the performance of the system against the gold standard. The objective was to find out if the synsets were distributed in such a way that the plausible interpretations appear at the top of the list and the incorrect ones at the bottom. The evaluation was done in terms of *mean average precision* (MAP) at top 30 synsets.

$$MAP = \frac{1}{M} \sum_{j=1}^M \frac{1}{N_j} \sum_{i=1}^{N_j} P_{ji}, \quad (6)$$

where M is the number of metonymic phrases, N_j is the number of correct interpretations for the metonymic phrase, P_{ji} is the precision at each correct interpretation (the number of correct interpretations among the top i ranks). First, the average precision was computed for each metonymic phrase independently. Then the mean values were calculated for the development and the test sets.

The reasoning behind computing MAP instead of precision at a fixed number of synsets (e.g. top 30) is that the number of correct interpretations varies dramatically for different metonymic phrases. MAP essentially evaluates how many good interpretations appear at the top of the list, which takes this variation into account.

7.3 Results

We compared the ranking obtained by applying Zipfian sense frequency distribution against that obtained by distributing probability mass over senses uniformly (baseline). We also considered the rankings before and after gloss processing. The results are shown in Table 6. These results demonstrate the positive contribution of both Zipfian distribution and gloss processing to the ranking.

7.4 Human Experiment

We conducted an experiment with humans in order to prove that this task is intuitive to people, i.e. they agree on the task.

We had 8 volunteer subjects altogether. All of

| Dataset | Verb Probability Mass Distribution | Gloss Processing | MAP |
|-----------------|---------------------------------------|---------------------|------|
| Development set | Uniform | No | 0.51 |
| Development set | Zipfian | No | 0.65 |
| Development set | Zipfian | Yes | 0.73 |
| Test set | Zipfian | Yes | 0.83 |

Table 6: Evaluation of the Model Ranking

| Group 1 | Group 2 |
|------------------|----------------|
| finish video | finish project |
| start experiment | begin theory |
| enjoy concert | start letter |

Table 7: Metonymic Phrases for Groups 1 and 2

them were native speakers of English and non-linguists. We divided them into 2 groups: 4 and 4. Subjects in each group annotated three metonymic phrases as shown in Table 7. They received written guidelines, which were the only source of information on the experiment.

For each metonymic phrase they were presented with a list of 30 possible interpretations produced by the system. For each synset in the list they had to decide whether it was a plausible interpretation of the metonymic phrase in an imaginary context.

We evaluated interannotator agreement in terms of Fleiss’ kappa (Fleiss, 1971) and f-measure computed pairwise and then averaged across the annotators. The agreement in group 1 was 0.76 (f-measure) and 0.56 (kappa); in group 2 0.68 (f-measure) and 0.51 (kappa). This yielded the average agreement of 0.72 (f-measure) and 0.53 (kappa).

8 Linguistic Perspective on Logical Metonymy

There has been debate in linguistics literature as whether it is the noun or the verb in the metonymic phrase that determines the interpretation. Some of the accounts along with our own analysis are presented below.

8.1 The Effect of the Noun Complement

The interpretation of logical metonymy is often explained by the lexical defaults associated with the noun complement in the metonymic phrase. Pustejovsky (1991) models these lexical defaults in the form of the *qualia structure* of the noun. The qualia structure of a noun specifies the following aspects of its meaning:

- **CONSTITUTIVE Role** (the relation between an object and its constituents)
- **FORMAL Role** (that which distinguishes the object within a larger domain)
- **TELIC Role** (purpose and function of the object)
- **AGENTIVE Role** (how the object came into being)

For the problem of logical metonymy the telic and agentive roles are of particular interest. For example, the noun *book* would have *read* specified as its telic role and *write* as its agentive role in its qualia structure. Following Pustejovsky (1991; 1995) and others, we take this information from the noun qualia to represent the default interpretations of metonymic constructions. Nevertheless, multiple telic and agentive roles can exist and be valid interpretations, which is supported by the evidence derived from the corpus (Verspoor, 1997).

Such lexical defaults operate with a lack of pragmatic information. In some cases, however, lexical defaults can be overridden by context. Consider the following example taken from Lascarides and Copestake (1995).

- (6) My goat eats anything. He really enjoyed your book.

Here it is clear that *the goat enjoyed eating the book* and not *reading the book*, which is enforced by the context. Thus, incorporating the context of the metonymic phrase into the model would be another interesting extension of our experiment.

8.2 The Effect of the Metonymic Verb

By analysing phrases from the dataset of Lapata and Lascarides (2003) we found that different metonymic verbs have different effect on the interpretation of logical metonymy. In this section we provide some criteria based on which one could classify metonymic verbs:

- **Control vs. raising.** Consider the phrase *expect poetry* taken from the dataset of Lapata and Lascarides. *Expect* is a typical object raising verb and, therefore, the most obvious interpretation of this phrase would be *expect someone to learn/recite poetry*, rather than *expect to hear poetry* or *expect to learn poetry*, as suggested by the model of Lapata

and Lascarides. Their model does not take into account raising syntactic frame and as such its interpretation of raising metonymic phrases will be based on the wrong kind of corpus evidence. Our expectation, however, is that control verbs tend to form logical metonymies more frequently. By analyzing the lists of control and raising verbs compiled by Boguraev and Briscoe (1987) we found evidence supporting this claim. Only 20% of raising verbs can form metonymic constructions (e.g. *expect, allow, command, request, require* etc.), while others can not (e.g. *appear, seem, consider* etc.). Due to both this and the fact that we build on the approach of Lapata and Lascarides (2003), we gave preference to control verbs to develop and test our system.

- **Activity vs. result.** Some metonymic verbs require the reconstructed event to be an *activity* (e.g. *begin writing the book*), while others require a *result* (e.g. *attempt to reach the peak*). This distinction potentially allows to rule out some incorrect interpretations, e.g. a resultative *find* for *enjoy book*, as *enjoy* requires an event of the type *activity*. Automating this would be an interesting route for extension of our experiment.
- **Telic vs. agentive vs. other events.** Another interesting observation we made captures the constraints that the metonymic verb imposes on the reconstructed event in terms of its function. While some metonymic verbs require rather *telic* events (e.g., *enjoy, want, try*), others have strong preference for *agentive* (e.g., *start*). However, for some categories of verbs it is hard to define a particular type of the event they require (e.g., *attempt the peak* should be interpreted as *attempt to reach the peak*, which is neither *telic* nor *agentive*).

9 Conclusions and Future Work

We presented a system producing disambiguated interpretations of logical metonymy with respect to word sense. Such representation is novel and it is intuitive to humans, as demonstrated by the human experiment. We also proposed a novel scheme for estimating the likelihood of a WordNet synset as a unit from a non-disambiguated corpus.

The obtained results demonstrate the effectiveness of our approach to deriving metonymic interpretations.

Along with this we provided criteria for discriminating between different metonymic verbs with respect to their effect on the interpretation of logical metonymy. Our empirical analysis has shown that control verbs tend to form logical metonymy more frequently than raising verbs, as well as that the former comply with the model of Lapata and Lascarides (2003), whereas the latter form logical metonymies based on a different syntactic frame. Incorporating such linguistic knowledge into the model would be an interesting extension of this experiment.

One of the motivations of the proposed sense-based representation is the fact that the interpretations of metonymic phrases tend to form coherent semantic classes (Pustejovsky, 1991; Pustejovsky, 1995; Godard and Jayez, 1993). The automatic discovery of such classes would require word sense disambiguation as an initial step. This is due to the fact that it is verb senses that form the classes rather than verb strings. Comparing the interpretations obtained for the phrase *finish video*, one can clearly distinguish between the meaning pertaining to the creation of the video, e.g., *film, shoot, take*, and those denoting using the video, e.g., *watch, view, see*. Discovering such classes using the existing verb clustering techniques is our next experiment.

Using sense-based interpretations of logical metonymy as opposed to ambiguous verbs could benefit other NLP applications that rely on disambiguated text (e.g. for the tasks of information retrieval (Voorhees, 1998) and question answering (Pasca and Harabagiu, 2001)).

Acknowledgements

I would like to thank Simone Teufel and Anna Korhonen for their valuable feedback on this project and my anonymous reviewers whose comments helped to improve the paper. I am also very grateful to Cambridge Overseas Trust who made this research possible by funding my studies.

References

- S. Abney. 1996. Partial parsing via finite-state cascades. In J. Carroll, editor, *Workshop on Robust Parsing*, pages 8–15, Prague.
- O. E. Andersen, J. Nioche, E. Briscoe, and J. Carroll. 2008. The BNC parsed with RASP4UIMA. In *Proceedings of the Sixth International Language Resources and Evaluation Conference (LREC'08)*, Marrakech, Morocco.
- B. Boguraev and E. Briscoe. 1987. Large lexicons for natural language processing: utilising the grammar coding system of the *Longman Dictionary of Contemporary English*. *Computational Linguistics*, 13(4):219–240.
- E. Briscoe, A. Copestake, and B. Boguraev. 1990. Enjoy the paper: lexical semantics via lexicology. In *Proceedings of the 13th International Conference on Computational Linguistics (COLING-90)*, pages 42–47, Helsinki.
- E. Briscoe, J. Carroll, and R. Watson. 2006. The second release of the rasp system. In *Proceedings of the COLING/ACL on Interactive presentation sessions*, pages 77–80.
- L. Burnard. 2007. *Reference Guide for the British National Corpus (XML Edition)*.
- D. Fass. 1991. met*: A method for discriminating metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.
- C. Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database (ISBN: 0-262-06197-X)*. MIT Press, first edition.
- J. L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.
- D. Godard and J. Jayez. 1993. Towards a proper treatment of coercion phenomena. In *Sixth Conference of the European Chapter of the ACL*, pages 168–177, Utrecht.
- M. Lapata and A. Lascarides. 2003. A probabilistic account of logical metonymy. *Computational Linguistics*, 29(2):261–315.
- A. Lascarides and A. Copestake. 1995. The pragmatics of word meaning. In *Journal of Linguistics*, pages 387–414.
- M. Pasca and S. Harabagiu. 2001. The informative role of WordNet in open-domain question answering. In *Proceedings of NAACL-01 Workshop on WordNet and Other Lexical Resources*, pages 138–143, Pittsburgh, PA.
- J. Preiss. 2006. Probabilistic word sense disambiguation analysis and techniques for combining knowledge sources. Technical report, Computer Laboratory, University of Cambridge.
- J. Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4).
- J. Pustejovsky. 1995. *The Generative Lexicon*. MIT Press, Cambridge, MA.
- M. Utiyama, M. Masaki, and I. Hitoshi. 2000. A statistical approach to the processing of metonymy. In *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany.
- C. M. Verspoor. 1997. Conventionality-governed logical metonymy. In *Proceedings of the Second International Workshop on Computational Semantics*, pages 300–312, Tilburg.
- E. M. Voorhees. 1998. Using WordNet for text retrieval. In C. Fellbaum, editor, *WordNet: An Electronic Lexical Database*, pages 285–303. MIT Press.