

Covert Events and Qualia Structures for German Verbs

Stefan Rüd (stefan.rued@gmx.de)

Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung
Azenbergstr. 12, D-70174 Stuttgart, Germany

Alessandra Zarccone (zarconaa@ims.uni-stuttgart.de)

Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung
Azenbergstr. 12, D-70174 Stuttgart, Germany

Abstract

Sentences like *The author began the book* (logical metonymies) involve the interpretation of covert events which are not explicitly realized on the surface (→ *The author began writing the book*). Qualia-based accounts of logical metonymies (Pustejovsky, 1991, 1995) account for such covert events using complex lexical entities (qualia structures) for the objects. We present a corpus study for the German verbs *geniessen*, *anfangen* (*mit*), *beginnen* (*mit*), *aufhören* (*mit*), based on data obtained from the deWaC corpus. In order to evaluate to what extent covert events in logical metonymies can be accounted for by qualia structures, instances of logical metonymies for these verbs were collected; paraphrases for the covert events were then manually annotated and compared with the qualia structures of the objects. We also analyzed sentences where the event was made explicit (long forms: *The author began writing the book*), comparing those events with the qualia events. We contrasted results for the two structures (metonymies - long forms) and across verbs, evaluating what sort of contribution qualia can make to logical metonymy resolution and what issues it poses.

Keywords: Logical metonymy; qualia; corpus study; German.

Introduction

Covert Events (CEs) are events which are not explicitly realized on the surface of a sentence, but play a key role in the understanding of some linguistic constructions. Classic examples are sentences like (1) or (2) - note that the CE can be formulated explicitly through an appropriate paraphrase (**long form**):

1. The author began the book → The author began writing the book
2. She enjoyed the film → She enjoyed watching the film

Typical situations for the occurrence of CEs are constructions (as these examples), where event-subcategorising verbs are combined with entity-denoting objects (**logical metonymies**). According to Pustejovsky's theoretical approach, a (semantic) type conflict triggers the recovering of CEs, which are in many cases derived from the so-called **Qualia Roles (QRs)**, contained in the qualia structures of the respective objects (Pustejovsky, 1995).

Qualia structures and Logical Metonymy

Qualia structures can be viewed as complex lexical entries, representing aspects of meaning of a word, its semantic relations and the roles involved in its understanding. They consist of four QRs, corresponding to four aspects of meaning (*constitutive quale*, *formal quale*, *agentive quale*, *telic quale*), although not all four need be present for each lexical unit. For

the scope of our study, only the agentive quale and the telic quale are relevant, because they are seen as those responsible for the understanding of logical metonymies (Pustejovsky, 1995; Briscoe, Copestake, & Boguraev, 1990):

agentive quale (AQ) contains information about the factors and causal chains involved in the coming about of the object (*book* → writing)

telic quale (TQ) denotes the purpose / function of an object (*book* → reading)

In logical metonymies, QRs of nouns make it possible to specify the semantics of the governing verbs by selecting an event from the QRs of the corresponding argument (Konerding, 2006):

3. Julia enjoyed the film [TQ = 'watch'; AQ = 'make']
4. Susanne began her first novel [TQ = 'read'; AQ = 'write']

In sentence 3, the VP *to watch the film* is formed by extending the meaning of the object *film* (type: entity) to a phrase (type: event) by integrating it with the missing information extracted from the QRs of the object (*to watch*, TQ of the object *film*). Which quale is used (AQ or TQ), depends primarily on the metonymic verb itself and on the QRs of the object, but often also on the context and on the subject.

Briscoe et al. (1990) and Verspoor (1997) observed that the QRs provide default CE interpretations, which can be overridden if a different interpretation is inferred from the context. For example, the lexicon entry for *to enjoy* should indicate by default that in cases of type coercion the specification of the CE comes from the QRs of the object (with a strong preference for its TQ). However, in cases where the default interpretation is overridden by contextual information, this leads to a pragmatically more appropriate interpretation. Lascarides and Copestake (1995), though arguing that the lexicon must contain generalizations, claimed that world knowledge (pragmatic knowledge) has priority over these general rules. If, as in the example 5, the object of *to enjoy* is an artifact, then the general rule is that the CE should be determined by the TQ of the object (*reading*).

5. The goat enjoyed the book. → eating
- 6.(a) The publisher began a series of books. → issuing
(b) The author began a series of books. → writing

However, the knowledge that goats do not usually read results in an exception to this rule and leads to the CE *eating* in 5. Even when different contexts suggest different predicates for paraphrasing CEs (examples 6a and 6b), even then they can both clearly be attributed to the AQ of *book* (see also Zarcone and Padó (2010)).

Qualia as prototypical concepts

For almost every noun there is a large number of verbs that can take it as an object, and all these verbs are in principle available for paraphrasing of CEs. For example, see the most frequent verbs for the object *Buch* (*book*):

Buch (101241): 5006 lesen, 3468 schreiben, 1561 geben, 1092 kaufen, 1018 veröffentlichen, 893 empfehlen, 619 machen, 581 finden, 566 nehmen, 464 legen, 435 vor#stellen, 385 kennen, 370 lassen, 357 bestellen, 350 finden, 331 verfassen, 326 machen, 301 verkaufen, 267 führen, 261 halten, 259 aus#leihen, 236 bringen, 233 erscheinen, 229 heraus#geben, 215 bekommen, 200 ab#runden, 186 sehen, 182 vor#legen, 176 heraus#bringen, 164 brauchen, 162 verschlingen, 161 auf#schlagen, 160 nennen, 154 durch#lesen, 151 erhalten, 149 schenken, 140 besitzen, 138 suchen, 126 publizieren, ...

If we assume that the QRs model conceptual knowledge, then the QRs can be considered prototypical concepts, bundling all these relations into four roles. If QRs were to be identified with single predicates, many metonymies could not be satisfyingly treated by qualia-based theories. If we now select the most common verbs for *Buch* which we could assign to the AQ and the TQ, the following picture emerges:

AQ: schreiben, veröffentlichen, machen, verfassen, vorlegen, publizieren, herausbringen, drucken, ...

TQ: lesen, durchlesen, studieren, ...

Depending on context or on the style, a variety of predicates can be assigned to each of the QRs, provided that these can be interpreted as instances of the corresponding concepts. On the other hand, some metonymies pose problems when looking for an appropriate paraphrase or choosing among different possible formulations, while still making sense to the speaker:

7. If the owner is not there, the employees can enjoy his boat.

While we are able to understand sentence 7, it would still be difficult to give a CE paraphrase for it.

Previous empirical studies

To what extent then can CEs be explained by a qualia-based theory? For the English language, Briscoe et al. (1990) and Verspoor (1997) carried out corpus-based studies for metonymic verbs. Briscoe et al. (1990) labelled as “pragmatic” those cases in which the CE does not arise from the QRs, but must be inferred from the context; their study on the LOB (Lancaster-Oslo/Bergen) corpus found that on average 17% of metonymies for the following verbs are pragmatic cases: *enjoy*, *prefer*, *finish*, *start*, *begin*, *miss* and *regret*. Verspoor (1997) found in a BNC corpus-based study that the CEs for the verbs *begin* and *finish* are determined by the QRs in about the 95% of the cases, and in about the 5% by the context - again, a prevalence for AQ or TQ interpretations.

The aim of our study

The aim of our study was to evaluate the role of QRs in the interpretation of logical metonymies by estimating to what extent the interpretation of instances of logical metonymy from a very large corpus can be accounted for by QRs. Compared to earlier studies, we provide an investigation for a different language (German), that is based on a much larger corpus (deWac, 1.7 billion words, compare with BNC, 100 million words, and LOB, a million words).

We present a study for four German verbs, based on data obtained from the deWac corpus: we harvested subject-verb-object combinations involving logical metonymies (**metonymic sentences** e.g. *Er fing einen Brief an* - *He began the letter*). CE paraphrases were then manually annotated, in the attempt to find the appropriate paraphrase in a specific context. The objects for the extracted sentences were then annotated with their AQ and TQ. From these annotations, we could determine the percentage of matches between the paraphrases and the QRs.

A second innovation of our study is an analysis of instances of **long forms**, where the event is made explicit with a dependent verb (e.g. *Er fing an, einen Brief zu lesen* - *He began reading the letter*), in order to compare them with the corresponding metonymies and to evaluate how close to the QRs such explicit events are.

Method

In order to facilitate comparisons with previous studies (Briscoe et al., 1990; Verspoor, 1997), the following German verbs were selected: *geniessen* (*to enjoy*), *anfangen* (*mit*) (*to start (with)*), *beginnen* (*mit*) (*to begin (with)*), *aufhören* (*mit*) (*to stop (with)*), and *beenden* (*to finish*). We used a dependency parsed version of the deWac corpus, a very large collection of German sentences of about 1.7 billion words, gathered from the Internet and made available by the WaCky project (Baroni, Bernardini, Ferraresi, & Zanchetta, 2008). The corpus was parsed with the BitPar parser (Schmid, 2004) and the FSPAR parser (Schiehlen, 2004). Since the number of occurrences of the selected verbs is very large in deWac corpus and only a relatively small number of them are examples of logical metonymy, we developed a Python script to heuristically select instances of logical metonymy. Our method consisted of the following steps:

- 1. Sentence extraction** - appropriate instances of metonymy (metonymic sentences) and corresponding long forms (where the event was explicit) were extracted;
- 2. Sentence annotation** - both metonymic sentences and long forms were annotated with regard to the type of implicit CE or explicit event, and (when possible) with the AQ and TQ of the corresponding object;
- 3. Evaluation of CE-QRs matching** - evaluation of the matches between CEs and QRs in the metonymic sentences and between the explicit events and the QRs in the long forms.

Table 1: Frequencies of the extracted sentences from the deWaC Corpus

verb	Occurrences (raw data)	+ dependent NP				+ dependent VP			
		total NP occur. (raw data)		from total NP artifacts		total VP occur. (raw data)		from total VP artifacts	
				(metonymic sentences)	events			(long forms)	events
geniessen	20749	20477	98.7%	5.5% (*)	6.4% (*)	272	1.3%	29.6% (*)	15.7% (*)
anfangen	5463	2571	47.1%	4.1% (*)	0.04% (*)	2892	52.9%	15.7% (*)	7.8% (*)
anfangen mit	4015	3691	91.9%	9.4% (*)	1.1% (*)	324	8.1%	14.5% (*)	10.5%
beginnen	41288	30111	72.9%	0.8% (*)	— (*)	11177	27.1%	9.8% (*)	9.7%
beginnen mit	36853	34858	94.6%	1.2% (*)	2.2%	1995	5.4%	5.2% (*)	9.2% (*)
aufhören	1223	13	1.1%	7.7% (*)	— (*)	1210	98.9%	8.2% (*)	8.7%
aufhören mit	1223	1188	97.1%	3.9% (*)	0.8% (*)	35	2.9%	14.3% (*)	13.8%
beenden	12014	12014	100.0%	2.0% (*)	0.02% (*)	—	—	—	—

Some cells are marked with asterisks (*): the corresponding sentence sets were manually checked for correct classification. Due to the large number instances, not all of them were individually checked.

1. Sentence extraction

We selected instances where both a subject and either an NP object (metonymic sentences) or a VP complement (long forms) were present, thus excluding a very large number of sentences where, for instance, the verb *begin* was used as intransitive (e.g. *Der Film begann - the movie began*). The drawbacks of this approach are (a) the loss of passives and (b) a reliance on accurate parsing results, increasing precision at the expense of recall; however, automatization of this step is crucial to analyse our large corpus. +NP and +VP instances are listed in Table 1.

Alongside metonymic sentences, long forms were also selected, where the event is explicit and has the form of a dependent verb (e.g. *Er fing an, einen Brief zu lesen - He began reading the letter*).

Resolution of underspecifications: Morphosyntactic information was used to distinguish subjects and objects; the FSPAR parse trees however allow for alternatives (e.g. underspecified case, 'Nom|Akk') for a large proportion of nouns. In such cases, if the script had not encountered a subject left to the verb yet, the underspecified noun was considered as a subject, otherwise as an object¹.

Selection based on semantic types: Since only artifacts have an AQ and a TQ and since CEs only occur if the subject is able to have intentions (and in the case of *enjoy* to have emotions, too), 'humans' and 'artifacts' are the most relevant categories for discovering CEs. We compiled lists of nouns for three semantic classes, namely humans, artifacts, and events, based on GermaNet 5.1 (Lemnitzer & Kunze, 2002). A large proportion of subjects and objects was automatically assigned categories during extraction, nouns not included in GermaNet were annotated manually. Also, the semantic categories helped to exclude a large number of non-metonymic sentences. For example *geniessen*, besides the reading "get pleasure from", has a very common reading of "have the benefit of", which

does not evoke metonymic interpretations (*die Partei genießt Vertrauen - The party enjoys trust*). Sentences with this non-metonymic reading were excluded, because their objects were not artifacts. Percentages of artifacts, events and other objects are listed in Table 1. Some cells are marked with asterisks (*): the corresponding sentence sets were manually checked for correct classification. Due to the large number instances, not all of them were individually checked.

2. Sentence annotation

Incorrectly extracted sentences were removed. Some errors were due to parsing errors in the sentence structure or in the subject annotation. Also, sentences with metaphoric readings were excluded, such as the word *Seiten* (pages) as an object of *geniessen* (enjoy) in *die schönen Seiten des Lebens* (the good sides of life). Also, many nouns were ambiguous between an entity and an event reading, e.g. *Malerei* (painting), *Bericht* (report), *Frühstück* (breakfast). Unless the context clearly suggested an artifact reading, these sentences were omitted. The proportion of discarded sentences for the above mentioned reasons was between 30% and 50%.

The selected metonymic sentences were then analyzed and annotated with CE paraphrases, and the QRs for their objects were determined. For long form sentences, the QRs for the objects were determined.

Context-dependent interpretation: In many cases, the paraphrase of a CE was so trivial that the explicit formulation would sound strange in a sentence, for example when the CE for objects of *geniessen* (to enjoy) refer to *eating* and *drinking*. On the other hand, many cases were not so trivial, due to lack of context; nevertheless, since the sentences were collected from the web, it was sometimes possible to find the original source, as in example 8:

8. *Wir haben mit einem traditionellem Brett angefangen und es lief recht gut.*
We started with a traditional board and it went quite well.

From the original website it was apparent here that a mother wants to teach her 8-years-old son to play chess.

¹In German (a relatively free constituent order language) the subject is still more likely to be left of the object.

Table 2: Annotator’s interpretations vs. Qualia interpretations (N.B. this table only refers to the artifacts column in Table 1)

verb	+ dependent NP (metonymic sentences)						+ dependent VP (long forms)			
	TQ	AQ	sum(AQ,TQ)	other	undetermined qualia	insufficient context	TQ	AQ	sum(AQ,TQ)	other
geniessen	89.7%	0%	89.7%	1.6%	6.9 %	1.9%	31.3%	28.1%	59.4%	40.6%
anfangen	21.7%	61.3%	83.0%	1.9%	3.8 %	11.3%	14.1%	24.0%	38.1%	61.8%
anfangen mit	33.7%	17.9%	51.6%	23.9%	1.7 %	22.8%	14.9%	0%	14.9%	85.1%
beginnen	5.6%	88.7%	94.3%	2%	0%	3.6%	12.1%	27.4%	39.5%	60.5%
beginnen mit	31.0%	35.4%	66.4%	20.3%	0%	13.3%	21.2%	0%	21.2%	78.8%
aufhören	100.0%	0%	0%	0%	0%	0%	33.0%	25.0%	58%	42.0%
aufhören mit	60.9%	23.9%	84.8%	6.5%	2.2%	6.5%	20.0%	0%	20.0%	80.0%
beenden	35.0%	49.8%	85.8%	8.4%	0%	6.8%	–	–	–	–

Insufficient context: If only little context was available, finding a suitable paraphrase was often not easy. For example, *Ich fange nochmals mit diesem Brief an* (*I’m starting again with this letter*) could be paraphrased with *to read*, *to write*, but also with *to talk about this letter*. In many cases “we start with X” means something like “we begin contemplating X” or “we begin enumerating X”, neither of which corresponds to the AQ or to the TQ. If only a short single sentence was available, it was often impossible to choose among several possible alternatives. Such instances were annotated with the label **insufficient context**.

PP attachment: *Anfangen* (*to begin*) without any object mostly means “to begin to work”, often with a PP as an adjunct. Therefore, with *anfangen mit* (*to start with*), *beginnen mit* (*to begin with*) and *aufhören mit* (*to stop with*), we must consider whether the PP is a verb argument, or whether it is only an adjunct:

9. *Er fängt mit dem Geschirr an.* → *Er fängt an, das Geschirr zu spülen.*
He begins with the dishes → He begins washing the dishes.
10. *Das Kind beginnt mit dem Ball* → *Das Kind beginnt mit dem Ball zu spielen.*
The child begins with the ball → the child begins to play with the ball .

In 9, *Geschirr* (*the dishes*) is the direct object of *spülen* (*washing*); however, in 10, the PP *mit dem Ball* is an oblique argument of *spielen* (*to play*) also in the paraphrase.

Undetermined qualia: For some general terms (e.g. *board*, *machine*) it was difficult to find an AQ or a TQ. Some of these cases could be solved by replacing the more general term with a more specific one (*chessboard*, *aircraft*), if suggested by the context (see also example 8), but some others (*city*, *garden*) were annotated as **qualia undetermined**.

Transparent nouns: In cases of transparent nouns such as *a cup of coffee* (Fillmore, Baker, & Sato, 2002), the content was regarded as the real object of interest (*coffee*), instead of the direct object of the verb (*cup*).

Specificity of Qualia: Often in Pustejovsky’s Generative Lexicon, it seems as if only two individual predicates can be assigned to the AQ and to the TQ respectively. In this study, it was assumed, in contrast, that the QRs are more likely to be understood as general concepts and therefore represented by a whole set of predicates. This was particularly problematic for long forms, when deciding whether a given verb is equivalent to a QR. Our strategy was rather generous, for example we considered *verschlingen* (*devour*) to be included in the TQ of *book*. Ultimately, the label **TQ** was given when the verb expressed a typical use of the object in the given context, and similarly the label **AQ** was given when the verb denoted a typical way to bring the object into being. For the verbs *kaufen* (*to buy*), *verkaufen* (*to sell*), *handeln mit* (*to trade with*), *sich beschäftigen mit* (*to deal with*), *experimentieren mit* (*to experiment with*), however, was consistently given the label **other**.

3. Evaluation of CE-QRs matching

We counted cases where a paraphrase for a CE matched (a) the AQ, (b) the TQ, or (c) neither (“other”), (d) cases where the qualia was undetermined (“undetermined quale”) or (e) the context was not sufficient to determine a paraphrase (“insufficient context”) - see table 2. For long forms, the match between the explicit event and the QRs for the objects were determined.

Results

Table 2 shows the relative frequencies (percentages) of CE paraphrases for the categories used: AQ, TQ, sum(AQ,TQ), other, undetermined qualia and insufficient context. Note that here we are only considering artifact objects both for metonymic sentences and for long forms.

Anfangen and *beginnen* yielded a similar profile: both “*anfangen* + direct object” and of “*beginnen* + direct object” have a strong preference for the AQ of the object. However, the corresponding combinations with *mit* (*anfangen mit* and *beginnen mit*) did not show this tendency (*beginnen mit* actually showed an opposite tendency towards the TQ).

In particular, two prominent groups of objects contributed to that high proportion of TQ for *anfangen mit*: 26 references

Table 3: Comparison with the Results of Briscoe et al. (1990) and Verspoor (1997)

Briscoe et al. (1990)									
verb	tot. occ.	tot. occurrences		NP		interpretation			
		tot. VP	tot. NP	Event	Entity			sum(AQ,TQ)	non-QRs
enjoy	65	6 (9.2%)	59 (90.8%)	21 (32.3%)	25 (38.5%)			21 (84%)	4 (16%)
start	136	73 (53.7%)	63 (46.3%)	42 (30.1%)	21 (15.4%)			21 (100%)	0 (0%)
begin	69	58 (84.1%)	11 (15.9%)	8 (11.6%)	3 (4.3%)			1 (33.3%)	2 (66.7%)
finish	39	8 (20.5%)	31 (79.5%)	8 (20.5%)	23 (59%)			17 (73.9%)	6 (26.1%)
Verspoor (1997)									
	tot. occ.		tot. NP		Entity	AQ	TQ	sum(AQ,TQ)	Context
begin	40407		4470 (11.1%)		164 (0.4%)	65 (39.6%)	91 (55.5%)	156 (95.1%)	8 (4.9%)
begin on					25	4 (16%)	5 (20%)	9 (36%)	16 (64%)
finish	11072		2799 (25.3%)		319 (2.9%)	94 (29.5%)	211 (66.1%)	305 (95.6%)	14 (4.4%)

were related to medications and drugs with the paraphrase *nehmen* (to take) (or *verabreichen*, to administer, if the subject is a doctor), 16 were related to the domain of feeding infants with several types of baby food (*Beikost*, *Milchbrei*, *Karottenbrei*) as objects.

Geniessen (to enjoy) is different from the first two verbs in many ways: a strong tendency for the TQ, a large number of trivial paraphrases among those TQ interpretations (the most frequent paraphrases are *to eat* and *to drink*), and a correspondingly low proportion of cases in which the context was not sufficient to determine a CE (1.9%). In contrast, the long form “*geniessen*(V(object))” occurred in very few instances. For example, *Film* occurred in more than 30 instances for *geniessen* as a direct object and in all these cases the CE was *to watch* (TQ), but only once with “*geniessen*(V(*Film*))”, contrastively referring to the AQ of the object:

11. *Ich habe es wirklich genossen, diesen Film zu drehen wenn man von den Szenen absieht, die ich bis zur Hüfte im Sumpf zubringen musste.*
I really enjoyed making this film apart from the scenes I had to spend up to the hip in the swamp.

The default interpretation of “*geniessen* + direct object” clearly corresponds to the TQ of the object, and it seems that the preferred way to express the AQ activity instead is to explicitly formulate it (as with *drehen* and *Film*). Similar patterns (TQ in the metonymic sentence, AQ or other in the long form) were found for *Haus* (house) and *Song* (song).

The verbs *aufhören* and *beenden* showed a complementary behavior: while *aufhören* has a very strong preference for VP-complements, *beenden* accepts only NP complements, and while metonymies with *aufhören* prefer the TQ, *beenden* shows a preference for the AQ.

The majority of interpretations for the metonymic sentences fall into the QR categories (sum(AQ,TQ)). More than 80% (and in some cases over 90%) of the CEs for *geniessen*, *anfangen*, *beginnen* und *beenden* correspond to the AQ or the TQ of the object, if this is an artifact. For *anfangen mit* and *beginnen mit*, this proportion is much lower (between 50% and 70%) and more than a fifth of the CEs in these cases differ from the AQ and from the TQ. In contrast, the

long forms, where the event is explicit, yielded a majority of context-based or “other” interpretations, which do not correspond with the QR events. This result is particularly interesting, because it confirms the observations in Lapata and Lascarides (2003) and Egg (2004) that metonymy is strongly related to Grice’s conversational maxims (Grice, 1975). If the QRs capture some basic/default event interpretation attached to the lexical representation of an artifact (e.g. *book* → *reading/writing*), we tend to omit that explicit information in a logical metonymy where the CE is retrievable from the QR (e.g. *John began the book* → *reading/writing*). If, on the other hand, the event is not a basic/default interpretation, but is a less typical one (e.g. *binding*), then we need to make it explicit in communication (e.g. *John began binding the book*).

Comparison with the results of Briscoe et al. (1990) and Verspoor (1997)

Table 3 provides a summary of the results of Briscoe et al. (1990) and Verspoor (1997) for the corresponding verbs to the German equivalents analyzed here. Their references were initially classified according to whether the complement of the verb was a VP or a NP. Furthermore, entity-denoting NPs that have metonymic interpretations were selected².

Briscoe et al. (1990) labelled as “pragmatic” those cases where the interpretation of the metonymy differed from the QRs of the object, but must be inferred from the context. We understand that this category must include what we called “other”, but not our “undetermined qualia” and “insufficient context” categories, since the authors report to have taken into consideration only those instances of logical metonymy where a paraphrase could clearly be determined. Unclear cases have been omitted in their study. Only two verbs (*begin* (on) and *finish*) were examined in Verspoor (1997) and a different reference value was used: the relative frequencies in percent refer to occurrences of the verb lemma in the corpus and not to the extracted examples (as we did).

²Note that our study only takes into consideration artifacts, whereas Briscoe et al. (1990) discuss *sea* and *river* have the TQ *to swim*, and Verspoor (1997) mentions *have family* and *do business* among her paraphrases for instances of metonymy.

Although any comparison between the results of such studies is problematic due to differences in corpus sizes and in the extraction methods, some interesting similarities emerge as well. Lexical differences were noted by all three studies. *Enjoy* matched the low non-QR interpretations of *geniessen*. *Beginnen* and *anfangen* have a strong AQ preference, while the English *begin* has more TQ interpretations. *Begin* and *begin on* in Verspoor (1997) respectively showed a preference for QR-interpretations and for context interpretations, and a similar contrast holds for our analysis of *anfangen (mit)*, *beginnen (mit)* and *aufhören (mit)*: the versions with *mit* have a significantly larger proportion of non-QR interpretations.

The general claim from Briscoe et al. (1990) and Verspoor (1997), that QRs can account for up to 80% of the reported metonymic instances, seems to be quite consistent with the results of the present work, which yielded values above 80% for *geniessen*, *anfangen*, *beginnen*, *aufhören* and *beenden*. For the combinations *anfangen mit*, *beginnen mit*, and *aufhören mit*, the number of non-QR interpretation is significantly higher than for the former mentioned verbs, though not as high as the value for *begin on* in Verspoor (1997).

Conclusions

We examined to what extent CEs in the metonymic use of some German verbs (*geniessen*, *anfangen (mit)*, *beginnen (mit)*, *aufhören (mit)*, *beenden*) can be explained by the QRs of the respective objects, using references extracted from the deWaC corpus. Our estimation was limited to artifacts, because they are the only entities that have an AQ and a TQ. For instances of logical metonymy, we estimated the frequencies for CE interpretations which could be accounted for by the QR of the object. For long forms, where the event was explicit, the subordinate verb was compared with the QR of the object.

CEs for *geniessen*, *anfangen*, *beginnen*, *aufhören*, *aufhören mit* and *beenden* are accountable for by either the AQ or the TQ of the object in more than 80% of the instances. This proportion was much lower for *anfangen mit* and *beginnen mit* (between 50% and 70%). More than a fifth of the CEs in these cases could not be recovered either from the AQ or the TQ. Also, different preferences among verbs for one or the other QR seem to emerge: *beginnen* and *anfangen* show a preference for the AQ, *geniessen* for the TQ.

In the long forms, a majority of these infinitives corresponds to non-QR interpretations, thus confirming the observations in Lapata and Lascarides (2003) and Egg (2004) that metonymy is strongly related to Grice's conversational maxims (Grice, 1975) (implicit CE, basic/default QR-interpretation; explicit event, non-QR interpretation).

A direct comparison between German and corresponding English verbs is difficult, also because the methods of selection and processing of the references are different. Nevertheless, some common patterns seemed to emerge: in particular, between *begin* and *begin on* in Verspoor (1997), and between our analysis of the constructions with *mit: begin on*

and the *mit*-constructions both have a significantly bigger proportion of non-QR interpretations than their versions without *on* and *mit*. The general claim from Briscoe et al. (1990) and Verspoor (1997), that QRs can solve up to 80% of the reported metonymic instances, was mirrored by our results, which include percentages of QR interpretations between 83% and 94% for *geniessen*, *anfangen*, *beginnen*, *beenden*.

References

- Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2008). The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation*, 43(3), 209–226.
- Briscoe, T., Copestake, A., & Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. In *Proceedings of the 13th COLING* (Vol. 2, pp. 42–47).
- Egg, M. (2004). Metonymie als phänomen der semantik-pragmatik-schnittstelle. *metaphorik.de*, 6, 37–53.
- Fillmore, C. J., Baker, C. F., & Sato, H. (2002). Seeing arguments through transparent structures. In *Proceedings of the third LREC* (pp. 787–791). Las Palmas, Spain: International Computer Science Institute.
- Grice, H. P. (1975). Logic and conversation. In P. C. et al. (Ed.), *Syntax and semantics* (Vol. 3). New York: Academic Press.
- Konerding, K.-P. (2006). Schichten, grenzen, gradationen. In E. W. Kirstel Proost (Ed.), *Von intentionalität zur bedeutung konventionalisierter zeichen, festschrift* (pp. 65–102). Tübingen: Institut fuer deutsche Sprache.
- Lapata, M., & Lascarides, A. (2003). A probabilistic account of logical metonymy. *Computational Linguistics*, 29.
- Lascarides, A., & Copestake, A. (1995). The pragmatics of word meaning. *Journal of Linguistics*, 387–414.
- Lemnitzer, L., & Kunze, C. (2002). Germanet – representation, visualization, application. In *Proceedings of the third LREC* (pp. 1485–1491).
- Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, 17(4), 409–441.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge: MIT Press.
- Schiehlen, M. (2004). Annotation strategies for probabilistic parsing in german. In *Proceedings of the 20th COLING* (pp. 390–396). Geneva, Switzerland.
- Schmid, H. (2004). Efficient parsing of highly ambiguous context-free grammars with bit vectors. In *Proceedings of the 20th COLING* (pp. 162–168). Geneva, Switzerland.
- Verspoor, C. M. (1997). *Contextually-dependent lexical semantics*. Doctoral dissertation, University of Edinburgh.
- Zarcone, A., & Padó, S. (2010). I like work: I can sit and look at it for hours - type clash vs. plausibility in covert event recovery. In *Proceedings of the VERB 2010 workshop*. Pisa, Italy.