Easy/*difficult*–constructions as triggers of implicit content:

comparing covert event elicitations and events extracted from a very large corpus

Alessandra Zarcone *, Janna Lipenkova °, Lukas Michelbacher *

* Institut für Maschinelle Sprachverarbeitung, ° Institut für Linguistik/Germanistik

* [zarconaa, michells]@ims.uni-stuttgart.de,
° janna.lipenkova@ling.uni-stuttgart.de

1 Introduction

Easy/*difficult*-constructions may involve the understanding of implicit events:

(1) a. *The book is difficult → to read / to write*
b. *The translation is easy*

The predicative adjective (*easy*/*difficult*) triggers an implicit event whose theme argument corresponds to the denotation of the subject NP in the source sentences (**covert event**, CE). We assume two possible sources for the elicitation of covert events, namely lexical knowledge and pragmatic knowledge (generalized event knowledge, see McRae & Matsuki 2009 for a review).

Lexical information is commonly seen as the "linguistic" part of knowledge involved in interpretation, in contrast to word knowledge. In terms of the Generative Lexicon (Pustejovsky 1995), the relevant part of lexical knowledge corresponds to the qualia structure, which defines the relational semantics of a lexical item; in particular, the source or the purpose of artifacts (e. g. *book*: *write* / *read*). In regards to formalization, qualia structures provide an economic and neat way to represent linguistic knowledge associated with lexical items.

Generalized event knowledge (that is, prototypical knowledge about typical events and typical scenarios where they take place) has been proven to be involved in online language processing to rapidly build effective expectation about plausible upcoming input. Generalized event knowledge originates from first and second-hand experience and is stored in our memory: for instance, we learn that people typically wash their hair indoor in a bathroom using shampoo, whereas they wash their cars in an outdoor environment not with shampoo, but with a hose. When hearing about somebody washing her/his car we might not know where the event take place, but we inevitably

build expectations about other elements involved in the scene (agents, patients, instruments, locations). These events cue elements of the scenario. Furthermore, it was also found that arguments easily recognized as taking part in a typical scenario can cue the corresponding events (McRae & Matsuki, 2009).

2 An experimental exploration

Speaker's judgments and corpus-extracted information are often used in experimental studies. Nevertheless, only rarely are such approaches combined, integrated and contrasted (Smith & Levy 2011). We present the first results from a two-fold study, aimed at exploring the extent to which corpus-extracted information can tap into covert events and generalized event knowledge.

The first part of the study elicited covert events for *easy*/*difficult* constructions. Elicitation is a major source for tapping into generalized event knowledge: related prototypical covert event information does not need to be explicit, but must be the first information coming to mind when speakers process sentences involving covert events:

(2) *The piano is difficult → to play*

The second part was a corpus extraction study. Extracting explicit realizations of events in *easy*/*difficult* constructions would have been problematic for several reasons. We wanted to avoid some circularity in our study: consider for example extracting "the piano is easy to X" and eliciting cloze completion data for the same construction. Also, contrast (2) with (3):

(3) *The piano is difficult to transport*

It seems unlikely that constructions where the event is explicit would indeed involve a typical event; it is more plausible to expect that, if the event is explicit, it is because the event does not denote a typical event (*to play* does not need to be explicit, *to transport* does), so typical events are by definition not attested in this form. Therefore, in order to avoid circularity and in order to obtain events that are indeed typical for a given object, we extracted all verbs having one of our object items as the head of their direct object.

2.1 Elicitation study

We elicited covert events for 60 sentences from 15 native speakers of English through a crowdsourcing platform for fast and affordable data collection (Amazon Mechanical Turk, Snow et al 2008). 30 objects belonging to 3 possible classes (10 per class) were used: entity-denoting objects (ENT, e.g. *the newspaper*), event-denoting objects (EVE, e.g. *the conference*) and entity/event ambiguous objects (AMB, e.g. *the breakfast*). The 30 objects were incorporated in stimuli sentences such as "the newspaper was difficult". Both *easy* and *difficult* were used, for a total

of 60 sentences, provided in randomized order. Participants were asked if the sentence involved "an additional activity that is not mentioned in the sentence" (e.g. *the newspaper was difficult to read*). If so, they were asked to provide one or more activities that could be plausible interpretations. Otherwise, they would tick "no additional activity".

Class	CE	no-CE
AMB	274	8
ENT	278	4
EVE	261	23

Table 1: CE and no-CE answers for the three classes of items

We analyzed both the binary answer "additional activity / no additional activity" (CE/no-CE) and the range of elicited activities. Participants were very productive: on average participants provided 4 different events for each sentence; across participants, events elicited per each noun ranged between 26 and 60 (mean 43). Event-denoting had the highest number of non-CE interpretations, but CE were still elicited in most cases (see Table 1). Pearson's chi-squared test (χ = 17.7353, df = 2, p-value < 0.001) confirmed that the counts for CE and no-CE answers were not proportional across the different classes of objects.

The lists of events per item were ranked with a mean reciprocal rank measure, in order to be compared with corpus-extracted list in the second part of the study. Table 2 lists the 10 top-ranked elicited events for three entity items, three event items and three ambiguous items. The events elicited can be classified into three groups: 1) "generic" events described by frequent verbs with very general meanings, e. g. *have*, *make*; 2) events corresponding to the qualia; 3) item-specific typical events.

We would expect to find the agentive (production) and telic (purpose) quale of those items, above all the entity items, but this does not appear to be entirely the case. Manufacturing an automobile is not part of our typical experience of automobiles (unless we are car manufacturers), more typical events are *buy, sell, fix, repair*. Letters on the other hand we do produce: *write* is one of the main events, together with *read*. Qualia events and typical events do not always overlap. Generic events are mostly the case for event-denoting items. Ambiguous items elicit events related to their entity component (*clean the shower*), but also to their event component (*finish the shower).*

2.2 Corpus study

In the second part of the study, lists of events were extracted from ukWaC, a large web corpus of English (2 billion tokens) created based on a web crawl of the .uk domain (Ferraresi et al. 2008). We used a version of ukWaC parsed with the Malt dependency parser (Nivre & Scholz 2004).

Given an object from the elicitation study, all the verbs having that item as the head of their direct object were extracted. The events in the lists for each object were ranked with regards to their frequency of co-occurrence with the object. Then, the events from the elicitation study were compared with those from the corpus study, in order to evaluate how well corpus-extracted events could predict the elicited events.

ENT	automobile	**elicitation: drive** (159), sell (78), fix (74), **buy** (67), repair (66), find (61), start (54), clean (37), steer (36), park (35)
		corpus: hire (69), schedule (61), **drive** (33), have (30), produce (22), make (21), build (19), purchase (15), use (14), **buy** (12)
	brandy	**elicitation: drink** (159), buy (56), taste (51), find (51), swallow (44), sell (39), **pour** (36), **make** (32), serve (27), identify (26)
		corpus: add (50), have (56), give (33), **drink** (28), **pour** (26), buy (16), sip (13), get (13), send (11), **make** (8)
	letter	**elicitation: write** (220), **read** (223), understand (72), **send** (72), mail (31), find (30), make (26), deliver (20), open (17), **receive** (14)
		corpus: write (12398), **send** (10468), **receive** (8890), have (3593), **read** (3013), get (2447), use (2188), see (1734), include (1414), cover (1254)
EVE	ceremony	**elicitation:** plan (91), **attend** (71), perform (61), hear (44), watch (36), understand (33), enjoy (22), sit through (21), get through (20), schedule (17)
		corpus: attend (1895), have (732), perform (708), hold (701), conduct (379), host (214), award (158), give (137), watch (136), arrange (117)
	debate	**elicitation:** win (91), hear (58), understand (49), listen to (46), attend (37), prepare for (35), participate in (31), follow (30), do (30), watch (29)
		corpus: have (3317), stimulate (1659), encourage (1140), open (943), inform (909), join (762), promote (639), hold (608), spark (594), lead (543)
	semester	**elicitation:** pass (87), **complete** (64), attend (51), finish (45), **take** (36), get through (27), endure (26), afford (25), enjoy (23), learn (23)
		corpus: spend (369), **take** (88), have (37), **complete** (37), study (34), teach (34), cover (28), last (24), run (24), hold (23)
AMB	breakfast	**elicitation: eat** (172), **cook** (124), **make** (111), prepare (71), digest (46), **serve** (38), clean up (28), finish (27), **enjoy** (26), order (24)
		corpus: have (2498), include (1348), **eat** (1097), **serve** (680), **enjoy** (670), **make** (392), get (351), offer (350), **cook** (324), provide (321)
	harvest	**elicitation: reap** (68), sell (68), **gather** (64), grow (51), harvest (36), plant (31), collect (30), take (29), tend (24), **get** (20)
		corpus: reap (283), produce (127), have (119), celebrate (84), ensure (66), bring (64), yield (60), see (57), **gather** (41), **get** (40)
	shower	**elicitation:** clean (83), **take** (69), finish (49), fix (36), plan (35), **install** (34), **enjoy** (28), start (27), prepare (20), run (19)
		corpus: have (1029), **take** (310), include (117), bring (80), provide (65), use (59), **install** (54), **enjoy** (54), dodge (54), get (52)

Table 2: three items for each class with their top five elicited events, together with their reciprocal rank measures, and their five most frequent events, together with their frequencies in ukWaC.

3 Implicit content

We used overlap coefficients as a similarity measure between a "gold standard" (elicited events) and a "test set" (corpus-extracted events) where the cardinality of the gold standard is smaller than that of the test set. An overlap coefficient is defined as the ratio between the cardinality of the intersection (gold ∩ test) and the cardinality of the gold standard. If the gold standard is a subset of the test set (best case scenario), then the overlap is equal to 1.

	E ∪ D	E ∩ D	Easy	Difficult
All	0.52	0.58	0.50	0.49
AMB	0.55	0.65	0.53	0.50
ENT	0.58	0.59	0.56	0.53
EVE	0.43	0.49	0.42	0.43

Table 3: overlap coefficients; gold standard: top 20 events elicited per item; test set: top 100 events from ukWaC

Table 3 reports the average overlap coefficients for the top 100 events for each item in the ukWaC evaluated against four different gold standards: the set "E ∪ D" (union set of events elicited for easy and events elicited for *difficult*), the set "E ∩ D" (events elicited both with *easy* and for *difficult*), the events elicited for *easy* and the events elicited for *difficult*.

Event-denoting had the highest overlap: typical events associated with entity-denoting items have a better chance of being retrieved from a corpus than typical events associated with event-denoting items. Pearson's chi-squared test ($\chi = 24.5908$, df = 2, p-value < 0.001) confirmed that the overlap counts were not proportional across the different classes of objects. The intersection most easily predicted by a corpus-extracted list, since its overlap coefficient is higher than the others ($\chi = 29.114$, df = 3, p-value < 0.001).

Corpus-extracted events do not show a striking overlap with elicited events. This improves when the intersection of events elicited for *easy* and events elicited for *difficult* is considered as the gold standard: events elicited for both adjectives are presumably less constrained by the lexical content of the adjectives and therefore easier to find in a corpus-extracted set of frequently co-occurring events.

3.1 *Easy*/*difficult*-constructions and logical metonymy

The concept of generalized event knowledge has been applied for event recovery in logical metonymies of the form NP V NP → NP V V NP (e. g. *The fan enjoyed the fight* → *watching the fight*) (Zarcone & Padó, 2011a, 2011b). In our results the available information appears less restrictive because no information about the agent is provided; logical metonymy yielded narrower results, which can be attributed to the biasing function of the agent information. This further supports the hypothesis that combinations of words jointly restrict the elicited event knowledge. Future studies modulating the effect of typical agents involving *easy*/*difficult* constructions are expected to yield similar results for such constructions as well.

4 Conclusions

The range of application of qualia structures is limited: it is defined only for physical artifacts and does not play into the interpretation of event-denoting objects. Nevertheless, participants also elicit covert events for event-denoting items, albeit to a lesser measure. Furthermore, the events recovered for entity-denoting items cover a wide range, often not overlapping with qualia events, but rather related to generalized knowledge about typical scenarios. These claims seem to point towards the hypothesis that covert events are retrieved from generalized event knowledge: the elicited information seems to be provided not by a constrained lexicon, but by a broader non-linguistic component of conceptual knowledge.

Also, the overlap between corpus-retrieved events and elicited events was not broad: the most prominent events in the elicitation task are typical in our everyday experience about objects, but they do not correspond in a straightforward manner to the events mostly associated with them in a large text corpus.

References

Ferraresi, A., E. Zanchetta, M. Baroni, & S. Bernardini (2008) Introducing and evaluating ukWaC, a very large web-derived corpus of English. In *Proceedings of the WAC4 Workshop at LREC 2008*.

Heim, I. (2000) Degree operators and scope. In: B. Jackson, T. Matthews, eds., *Proceedings of SALT 10*. CLC Publications, Ithaca, NY, 40–64.

McRae, K. & K. Matsuki (2009) People Use their Knowledge of Common Events to Understand Language, and Do So as Quickly as Possible. *Language and Linguistics Compass* 3: 1417-1429.

Nivre, J. & M. Scholz (2004) Deterministic Dependency Parsing of English Text. In *Proceedings of COLING 2004*, Geneva, Switzerland, August 23-27, 2004.

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

Smith, N. J. & R. Levy (2011). Cloze but no cigar: The complex relationship between cloze, corpus, and subjective probabilities in language processing. *Proceedings of CogSci 2011.*

Snow, R., B. O'Connor, D. Jurafsky, & A. Y. Ng (2008) Cheap and fast - but is it good? Evaluating nonexpert annotations for natural language tasks. In *Proceedings of EMNLP 2008*.

Zarcone, A. & S. Padó (2011a) "I like work: I can sit and look at it for hours" - Type clash vs. plausibility in covert event recovery". *Proceedings of Verb 2010*.

Zarcone, A. & S. Padó (2011b) Generalized Event Knowledge in Logical Metonymy Resolution. *Proceedings of CogSci 2011.*