

Easy/difficult–constructions as triggers of implicit content: comparing covert event elicitations and events extracted from a very large corpus

Alessandra Zarcone¹, Janna Lipenkova² and Lukas Michelbacher¹
 alessandra.zarcone@ims.uni-stuttgart.de, janna.lipenkova@ling.uni-stuttgart.de, lukas.michelbacher@ims.uni-stuttgart.de
¹ Institut für Maschinelle Sprachverarbeitung, ² Institut für Linguistik / Germanistik, Stuttgart, Germany

1. Easy/difficult-constructions

- * *The swim is difficult*
- * *The piano is difficult to play*
- * *The translation is easy*

$$[[difficult]] = \lambda e \lambda P \lambda x. difficult(e) \wedge P(x) = e$$

- * Easy/difficult subcategorize for events
- * Type-clash: triggers a **covert event (CE)**
- * How is this implicit knowledge retrieved?

	ENT		EVE		AMB	
	automobile	letter	ceremony	debate	breakfast	shower
elicitation	drive (159) sell (78) fix (74) buy (67) repair (66)	write (220) read (223) understand (72) send (72) mail (31)	plan (91) attend (71) perform (61) hear (44) watch (36)	win (91) hear (58) understand (49) listen to (46) attend (37)	eat (172) cook (124) make (111) prepare (71) digest (46)	clean (83) take (69) finish (49) fix (36) plan (35)
corpus	hire (69) schedule (61) drive (33) have (30) produce (22)	write (12398) send (10468) receive (8890) have (3593) read (3013)	attend (1895) have (732) perform (708) hold (701) conduct (379)	have (3317) stimulate (1659) encourage (1140) open (943) inform (909)	have (2498) include (1348) eat (1097) serve (680) enjoy (670)	have (1029) take (310) include (117) bring (80) provide (65)

boldface: the event appears in both sets; underlined: the event is part of the qualia structure of the item

2. Coercion: a supertype of phenomena

- * Type clashes require to “fill in” the missing information
- * Type coercion: semantic operation that converts argument to the type that is expected by a function

- * *John began the beer* → *drinking the beer*
- * *The fast typist* → *the typist who types fast*

3. Lexicon vs. world knowledge

The lexical hypothesis (Pustejovsky 1995):

- * qualia structure in the lexicon (book: reading OR writing) => Cfr GL
- * economical, neat way to represent linguistic knowledge associated with lexical items
- * too restrictive: it only applies to artifacts

Generalized event knowledge (GEK) (McRae and Matsuki 2009):

- * prototypical knowledge about typical events and their participants (first and second-hand experience, available in our memory)

- * *wash hair* ⇨ shampoo, sink, bathroom, indoor
- * *wash car* ⇨ hose, outdoor

- * words in isolation immediately activate GEK
- * words can rapidly combine to cue specific concepts that are relevant to GEK scenarios

4. Research questions

- * can corpus-extracted typical events predict covert events elicited for easy/difficult-constructions?
- * can a qualia-based theory account for covert event retrieval in easy/difficult-constructions?

Two-fold exploration:

- * elicitation study
- * corpus extraction

5. Elicitation study

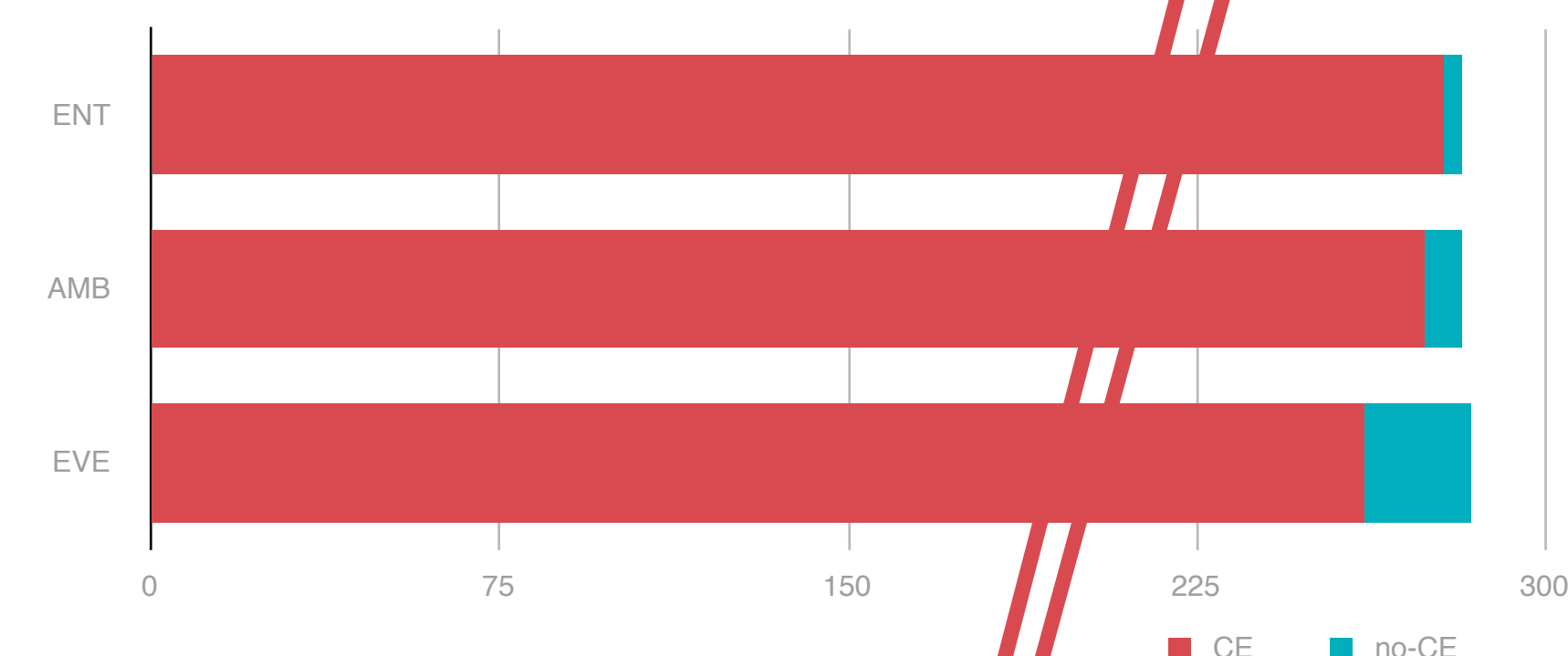
Materials:

- * 30 objects (10 x 3 classes):
 - ENT (entity-denoting): *the newspaper*
 - EVE (event-denoting): *the conference*
 - AMB (entity/event-denoting): *the breakfast*
- * 30 objects x 2 adjectives (*easy / difficult*) = 60 stimuli sentences

Method:

- * 15 native speakers of English
- * crowdsourcing platform (Snow et al. 2008)
- * “*The newspaper was difficult*”
 - ➔ Does it involve an additional activity that is not mentioned in the sentence? (CE / no-CE, binary answer)
 - ➔ If it does, cloze completion task (covert event elicitation)

6. Binary answer: CE vs. no-CE



CE / no-CE answers for the three object classes
 ➔ Effect of obj. class on CE/no-CE counts
 $(\chi = 17.7353, df = 2, p\text{-value} < 0.001)$

7. Elicited CEs

ENT:

- * *letter*: write, read
- * *automobile*: telic quale (driving) but not agentive quale (produce), more typical events are buy, sell, fix, repair

EVE:

- * *difficult/easy* restrict the range of events to those for which the degree of difficulty is relevant (no light verbs)

AMB:

- * events related to their entity component (*clean the shower*), but also to their event component (*take the shower*)

8. Corpus extraction

Extraction of all verbs having one of the 30 object items as the head of their direct obj

Corpus:

ukWaC (2 billion token corpus of web English, Ferraresi et al. 2008), parsed with the Malt dependency parser (Nivre & Scholz 2004)

Problems with corpus-extracted events

- * light verbs (*take a shower*)
- * non discriminative verbs (*have breakfast*)
- * idiosyncrasies (*includes breakfast*)

9. Elicited events and corpus events

- * Elicited CEs ranked (mean reciprocal rank measure)
- * Corpus events ranked (obj-V co-occurrences)
- * Overlap measure between top 20 elicited events and top 100 corpus events

$$overlap(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)}$$

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

- * Grice’s Maxim of Quantity:

“Do not make your contribution more informative than is required”

10. Conclusions and future work

- * Lexical Hyp.: qualia are often a subset of elicited events, but sometimes are not elicited
- * GEK Hyp.: typical events elicited, but not frequent in the corpus
- * Future work:
 - * filter light verbs
 - * try to identify rare realizations of typical events in corpus data with association measures (Evert 2005)
 - * comparison with other types of coercion

Bibliography

Evert, S. (2005). The Statistics of Word Cooccurrences: Word Pairs and Collocations. Dissertation, Institut für maschinelle Sprachverarbeitung, University of Stuttgart.
 Ferraresi, A., E. Zanchetta, M. Baroni, & S. Bernardini (2008) Introducing and evaluating uk-WaC, a very large web-derived corpus of English. In Proceedings of the WAC4 Workshop at LREC 2008.
 Grice, Paul (1975). "Logic and conversation". In Syntax and Semantics, 3: Speech Acts, ed. P. Cole & J. Morgan. New York: Academic Press.
 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.

The research for this paper has been funded by the German Research Foundation (DFG) as part of the Graduate school of the SFB 732 at the University of Stuttgart.