A corpus-based study of (non-)exhaustivity in *wh*-questions

A key issue in *wh*-question interpretation regards the distribution of exhaustive (Mention-All, MA) vs. non-exhaustive (Mention-Some, MS) question readings (see (1) and (2)):

(1) Who came to the party?
    a. Who is every person that...?   MA
    b. Who is a person that...?   ?MS

(2) Where can I find coffee?
    a. What is every place that...?   MA
    b. What is a place that...?   MS

Linguists' intuitions have typically concluded that MA is generally appropriate, while MS is marked [1-10]. Linguistic factors have been noted to generate variation in readings, including the specific *wh*-word—e.g., *who*-questions are biased for MA, while *where*/*how*-questions are biased for MS [11-12]—and existential (priority) modality—e.g., *can* purportedly licenses MS, as in (2) [3-5,7-8]. Recent work [13] tested these judgements in lab-controlled experiments with artificial stimuli and found evidence for some biases; however, [13] showed these biases can be overridden by features of the context like speaker/discourse goals [2,7,10-12]. To-date there is no systematic investigation of *naturally occurring questions* that tests the intuitions reported in the literature. We ask: (Q1) How much does question interpretation vary in natural discourse contexts? Is there indeed a bias for MA? (Q2) Is the distribution of interpretations modulated by linguistic form? We addressed these questions in a two-part study.

Methods. **Step 1: Naturalistic Stimuli from a Corpus Database.** Using TGrep2 and the Tgrep2 Database Tools [14-16], we extracted all occurrences of *wh*-questions (10,009) from the Switchboard corpus [17] and coded the questions for syntactic structure (e.g., embedded, root), *wh*-word, and presence of modality. To curate stimuli for step 2, we focus on root and embedded questions, leaving 2070 unique *wh*-questions. The distribution of *wh*-word and modality is reported in Table 2. **Step 2: Paraphrase Rating Task.** The remaining cases were divided into 31 lists with occurrence of critical factors roughly proportional to the overall database. Participants (n=1740) on Prolific were presented with each question and the 10 preceding lines of dialogue, and asked to rate the likely intended meanings (paraphrases), using a slider task (Fig. 1). Question paraphrases were selected to reflect MS/MA readings: *a* indicates MS ((1b)/(2b)), *every* MA ((1a)/(2a)), while the two readings converge in *the*-paraphrase (*what/who is the place/person*). There was a fourth option (*something else*) in case no other was appropriate.

Results. Questions with highest ratings for *something else* (17%) were excluded because they were rhetorical (see Tab. 1). *The*-paraphrases, where MS=MA, had the highest mean rating (.55), suggesting that only one reading was possible for most cases. Data were analysed using linear mixed effects regression. To investigate the posited MA bias, we compared *every* vs. *a* ratings, as these represent MA and MS (Fig. 2): although there was no bias for *every* contrary to literature (Q1), means for MA were higher than MS agregating over root and embedded questions. However, significant 3-way interactions between paraphrase and linguistic form factors partially support reports from the literature (Q2). First, the presence of a modal resulted in higher ratings of *a* [5-9,10] but not *every* for all except for *where*-questions. Second, *how*,*why*, and *when*-questions all showed a bias for MS, confirming [3-4, 10]; *who* and *where* show no bias (except for 'the') in contrast, and finally *what* questions revealed a bias shifting from MA to MS with a modal present.

Conclusion. In contrast to theoretical predictions, we find no bias for MA question readings in naturalistic dialogue (Q1). With respect to (Q2), we find support for some, but not all, observations about the effect of linguistic form on question interpretation reported in the literature. We suggest that MS/MA readings result from reasoning about the speaker's goal in the context, consistent with a constraint-based account [18] on which hearers integrate multiple sources of information to determine meaning. These results also highlight the importance of large-scale experiments for insight into more realistic meaning distributions [19].

Paraphrase	Example	Mean
every (MA)	*Where have you skied?*	.66
	Where's it all going?	.59
a (MS)	*Where do you like to eat?*	.57
	How would you achieve that?	.51
the (MS=MA)	*Where you going to school?*	.99
	Where do you work?	.99
something else	*Who knows?*	.61
	How can you watch that?	.53

Table 1: For each paraphrase, examples of questions that resulted in high ratings on that paraphrase.

Wh-word	+Modal	-Modal
What	5.3%	41.28%
How	3.93%	23.38%
Where	1.72%	9.38%
Why	1.07%	4.69%
Who	0.35%	5.00%
When	0.24%	1.45%

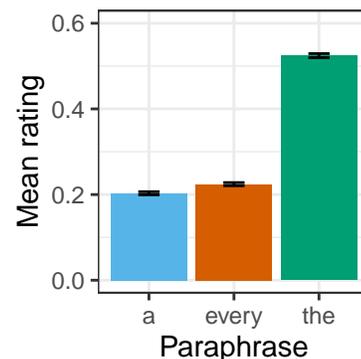**Table 2:** Distribution of *wh*-words and modality in Switchboard root and embedded questions. % of total (2070).

Figure 1: Paraphrase Rating Task: Participants evaluate intended question meanings by moving the slider next to paraphrases, assigning a numerical value between 0-1 to generate a proper probability distribution. Combined ratings must sum to 1.

References. [1] Karttunen (1977), [2] Groenendijk & Stokhof (1984), [3] George (2011), [4] Nicolae (2013), [5] Fox (2014, 2018), [6] Chierchia & Caponigro (2013) [7] Dayal (2016), [8] Xiang (2016, 2020), [10] van Rooij (2003), [11] Ginzburg (1995), [12] Asher & Lascarides (1998), [13] Moyer & Syrett (2019), [14] Rohde (2005), [15] Jaeger (2006), [16] Degen & Jaeger (2011), [17] Godfrey et al. (1992), [18] Degen & Tanenhaus (2019), [19] Degen (2015)

Figure 2: *Every*-paraphrases were not preferred over *a*-paraphrases.
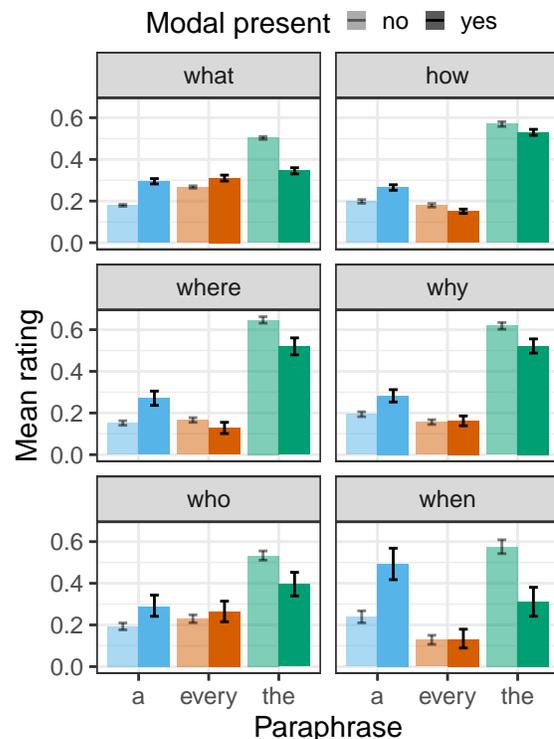
Figure 3: Significant 3-way interactions confirm some but not all intuitions from literature about linguistic form factors.