Exhaustive interpretations: what to say and what not to say

Benjamin Spector – Laboratoire de Linguistique Formelle, Paris 7/Ecole Normale Supérieure
(benjamin.spector@ens.fr)

I. When do answers give rise to an exhaustivity effect ? Suppose (1) is uttered in a context in which it is known that every person who came to a certain party belongs either to a certain group of philosophers, a certain group of chemists or a certain group of linguists:

(1) Who came to the party ?

Consider now the following answers that could be given in this context:

(2) a. Three philosophers. b. Some philosophers, and many chemists

(3) a. No philosopher b. Few philosophers, if any.

(4) a. Between 3 and 5 chemists b. Between 3 and 5 chemists and many linguists

(5) Some philosophers, and no chemists

Answers such as those given in (2), among others, are interpreted as *exhaustive*: one infers from (2)a that exactly three philosophers came to the party, and that there weren't any linguist nor any chemist, and (2)b. yields the inference that some but not many of the philosophers came, many but not all of the chemists came, and no linguist came. What the answers in (2) have in common is that they are *positive* answers (they consist of monotone increasing quantifiers). As shown by Groenendijk & Stokhof (1984), one can capture the generalization that positive answers give rise to such inferences by defining an exhausitivity operator that can be applied to positive answers and returns the desired readings. If P is the question predicate (in (1), the question predicate is "came to the party"), then the operator exh_P, which denotes a function from propositions (i.e. sets of worlds) to propositions, is defined as follows[1] (I am in fact using the definition proposed by Van Rooy & Schulz 2004, rather than the original formulation by Groenendijk & Stokhof –hereafter G&S):

(6) $[[exh_P\ S]] = \{w \in [[S]] :$ there is no $w' \in [[S]]$ such that $P(w') \subset P(w)$ and such that for every predicate Q distinct from P, $Q(w') = Q(w)\}$[2]

Yet things become more complex when one turns to non-positive answers, such as those exemplified from (3) to (5). If we were to apply the exhaustivity operator to sentences in (3) (i.e term-answers consisting of a decreasing quantifier), we would simply end up with the contradictory proposition. In fact, hearers tend to infer from such sentences either that non-philosophers are taken to be "irrelevant" by the speaker, or that the speaker has simply no idea as to which non-philosophers came (some informants say that they infer that there must have been a relatively important number of non-philosophers, but no one seems to infer that *all* the non-philosophers came). What about non-monotonic answers ((4) and (5))? (4)a and (4)b tend to trigger the inferences in (7)a and (7)b, respectively:

(7) a. Between 3 and 5 chemists came, but the speaker does not know the exact number, no philosopher came, no linguist came

b. Between 3 and 5 chemists came, but the speaker does not know the exact number, many linguists but not all came, and no philosopher came.

It thus turns out that the answers in (4) are interpreted as exhaustive, in the sense that they trigger a negative inference with respect to individuals that they do not "talk about" explicitly. Yet note that the exhaustivity operator would yield too strong results for such answers, since if *exh* were applied to them, they would be interpreted as implying that exactly three chemists came, contrary to fact (since one rather infers that the speaker does not know exactly how many

[1] Term-answers are analyzed as elided structures, i.e. an answer like "Three philosophers" is treated, if intended as an answer to (1), as identical to "Three philosophers came to the party"

[2] Notation: For any world w, 'P(w)' represents the extension of P in w; '⊂' represents *proper* inclusion.

chemists came). Moreover, one should not conclude that exhaustivity effects are found with all non-negative answers (i.e. positive answers and non-monotonic ones). Indeed (5) is non-monotonic too, but does not trigger an exhaustivity effect, since one does not infer from (5) that no linguist came. Rather, (5) triggers the inference that the speaker does not know much about linguists, or that, for some reason, he considers linguists to be irrelevant. These observations immediately raise a number of questions: a) How should we define exhaustivity in the general case, given that the exhaustivity operator yields too strong results when applied to some sentences which, intuitively, do trigger an exhaustivity effect? b) What are the right descriptive generalizations ? We have just seen that a taxonomy that would divide answers into three classes (positive, negative, non-monotonic) is too gross, since we need to make distinctions among the set of non-monotonic answers. c) Assuming that b. is answered, which general principles of conversational rationality could account for the generalizations we observe ?

II. Exhaustivity and the semantics of wh-questions. Recent works (Spector 2003, Van Rooy & Schulz 2004) have shown that a) the exhaustivity operator is able to account for so-called *scalar implicatures* and b) exhaustive interpretations can be derived from general pragmatic principles, in particular from some version of Grice's maxim of quantity. The precise derivation of exhaustive interpretations relies on the following assumptions:

a) Maxim of *positive quantity*: the speaker's answer must express the most informative *positive proposition* that he deems true, where a proposition S is defined as positive, relatively to a question of the form "?xP(x)" ("Which objects have property P?") if S belongs to the closure under disjunction and conjunction of the set of propositions that can be expressed by a sentence of the form 'P(c)' (assuming that each individual in the contextually given domain of quantification is named by some constant). b) *Competence assumption*: the speaker is as informed as possible relatively to the question under discussion, given that he has complied with conversational maxims. It can indeed be proved that a) and b), together with the assumption that the speaker believes his answer to be true (maxim of *quality*), entail that the speaker who uttered a positive answer is in an information state that entails the exhaustive reading of his answer.

An unmotivated aspect of this account is the specific role played by the notion of *positivity*. According to G&S's partition semantics for questions, which is the framework within which the above-mentioned accounts have been developed, a question of the form "?xP(x)" denotes a function that maps every world w to the set of worlds in which the extension of P is the same as in w (i.e. to the proposition that expresses the *complete answer* to the question in world w). It follows that a speaker who knows, say, that John, and nobody else, came to the party should say *explicitly* that nobody else came (rather than simply say "John came", as is in fact typically the case). More generally, we expect the maxim of quantity to require that speakers utter an answer that expresses explicitly all the *relevant* information they have (G&S in fact explicitly modeled the maxim of quantity in these terms), rather than simply their *positive* information. The very notion of *positivity* can actually not be expressed solely in terms of the question's denotation as G&S views it, since, once the domain of quantification is fixed, two questions such as "who came?" and "who didn't come?" have in fact the same denotation according to partition semantics (yet a positive answer to the first one is a negative answer to the second one). In order to overcome this difficulty, I will provide a new semantics for wh-questions, according to which a question of the form "?xP(x)" denotes a function that maps every world w to the *most informative positive proposition* that is true in w (where the definition of positivity is relative to the predicate P). I will discuss the relationship between this revised semantics and partition semantics, on the one hand, and Hamblin (1973) and Kartunnen

(1979), on the other hand; I will show that the specific insights of these different approaches can be captured within the new proposal; I will also discuss the interpretation of embedded questions (partly along the lines of Heim 1994).

III. The asymmetry between positive and negative information. Once a question is seen as requiring that speakers give all their positive information, the asymmetry between positive and negative answers is expected: a speaker who chooses a purely negative answer thereby indicates that he doesn't have any positive information, which explains why negative answers do not trigger positive inferences with respect to individuals they do not talk about (cf. (3)), contrary to what von Stechow & Zimmerman (1984) claims. But a negative answer also generally triggers the inference that the speaker has no more negative belief that what is explicitly conveyed in the answer. This suggests that the maxim of quantity should also say something about negative information. For instance, it could state that *if an answer is negative (i.e. is equivalent to an element of the closure under conjunction and disjunction of the negations of elementary answers), then it must express the strongest negative proposition that the speaker believes to be true* (*negative quantity, 1st version)*. If we extend this maxim so as to apply to all non-positive answers, to the effect that all non-positive answers must entail the strongest negative proposition that the speaker deems true, we predict that answers that are neither positive nor negative such as those in (4) and (5), do not trigger any exhaustivity effect: indeed authors of such sentences are then supposed to have provided both their positive and negative information explicitly. However, as mentioned above, this is not always so. Consider again (4)a.: on the one hand, we do not want to assume that the speaker has provided all the negative information she has, since this would block the inference that the speaker believes that no non-chemist came; on the other hand, we infer that the speaker does not know the exact number of chemists who came, presumably because if she had known that no more that four came, for instance, she would not have said "between three and five chemists", i.e. one assumes that the speaker has at least given all the negative information she has *regarding chemists.* Suppose we modify *negative quantity* along the following lines:

Negative quantity (2nd version - informally): the speaker's answer must provide all the negative information the speaker has regarding the individuals that the answer, so to speak, "mentions negatively".

Positive quantity and *negative quantity* so defined, together with the competence assumption, correctly predict that a speaker who uttered (4)a. does not know how many chemists came, and believes that no non-chemist came (if the speaker is taken to be competent, one attributes to him as much negative information as is compatible with the assumption that he complied with conversational maxims). But these assumptions, if not supplemented by some other principle, wrongly predict that (5) triggers the inference that no linguist came. Something more is needed.

Consider the following constrast:

(8) a. Among Jack, Peter, and Mary, who came to the party ?
b. Peter   c. Peter, and not Mary

While one infers from (8)b. that neither Jack nor Mary came, one does not infer anything special about Jack when hearing (8)c (apart from the fact that probably the speaker does not know much about Jack). Several informants typically comment on this fact by saying : "if you utter c., then you must not know anything negative regarding Jack, for otherwise why would you have mentioned Mary but not Jack ?". I will argue that conversational maxims must be enriched with a principle according to which speakers cannot treat differently individuals with respect to which they are in the same kind of epistemic state. While this formulation is quite

vague, it turns out that what is actually needed to account for the facts boils down to the following principle:

Principle of epistemic symmetry: if the speaker's information state contains only negative information with respect to two individuals d and d', then his answer must treat them on a par, i.e. either mention negatively both of them, or not mention them at all.

Before formally defining the relevant notions, let me informally show how *positive quantity*, *negative quantity* (second version) and the principle of epistemic symmetry, together with the competence assumption, make the right predictions: (4)b is such that there is no individual d that is mentioned only negatively (intuitively, it provides both negative and positive information regarding chemists, and no negative information at all with respect to non-chemists); as a result, a speaker who believes that no philosopher came can have complied with the principle of epistemic symmetry by uttering (4)b. The competence assumption then leads to the conclusion that the speaker actually believes that no philosopher came. On the other hand, (5) mentions chemists only negatively, and does not negatively mention any non-chemist; therefore, a speaker who believes that no linguist came cannot utter (5) and still have complied with the principle.

Negative quantity and the *principle of epistemic quantity* cannot be formally expressed unless we specify what it means for an answer to be "negatively" or "positively" about some individual d. I therefore will use the following definitions: Let D be the contextually given domain of quantification. Let P be the question predicate and d a member of D. Then:

- A proposition S *positively P-concerns* d if there is a world w such that S is true in w and d belongs to the extension of P in w, and such that S is false in the world w' identical to w except that d does not belong to P in w' (i.e. the extension of P in w' is the same as in w minus {d}).
- A proposition S *P-negatively P-concerns* d if there is a world w such that S is true in w and d does not belong to the extension of P in w, and such that S is false in the world w' identical to w except that d does belong to P in w'.

It can be shown that, as long as D is finite, a proposition is positive (relatively to a question whose predicate is P) if and only if it does not positively P-concern any individual d.

The final versions of *negative quantity* and the *principle of epistemic symmetry* are as follows:

Negative quantity (final version): Let $E = \{e_1, \ldots, e_n\}$ be the set of elements of D that an answer A negatively P-concerns. Then A must entail the strongest E-negative proposition that the speaker deems true, where an E-negative proposition is a proposition that is equivalent to a member of the closure under conjunction and disjunction of the set of propositions of the form '$\neg P(e)$', where *e* is a rigid constant that refers to a member of E, and assuming that each member of E is named by some constant (this definition can in fact be expressed in purely semantic terms).

Principle of epistemic symmetry (final version): if the speaker's information state negatively P-concerns two individuals d and d' and does not positively P-concern neither d nor d', then his answer A must either negatively P-concern both d and d', or not P-concern them at all.

I then show that these three principles (*positive quantity*, *negative quantity*, *principle of epistemic symmetry*), together with the competence assumption, predict that an answer A to a question whose predicate is P gives rise to an exhaustivity effect (i.e. a negative inference with respect to the individuals that A does not P-concern) if and only if A is *quasi-positive*, in the following sense: A is *quasi-P-positive* if whenever A negatively P-concerns an individual d, A also positively P-concerns d.

As an illustration, note that the answers in (4) are quasi positive, but that (5) isn't.