From Inference to Meaning. Experimental Study on Reasoning with Quantifiers *Some* and *Most*.

Maria Spychalska

Institute of Philosophy, University of Warsaw
m.spychalska@gmail.com

Abstract. We report on the results of our reasoning experiments concerning direct inferences with quantifiers: *some*, *most*, *all*. We investigated scalar implicatures of *some* and *most*, as well as inferences from *all* to *most*, from *all* to *some* and from *most* to *some*. Based on our results, we propose that scalar implicatures are context-independent and default in this sense that the pragmatic interpretation of the lexical item with which they are connected is preferred in communication. Following Mostowski and Wojtyniak (2004), we observe that meaning of a sentence may be established in two ways: via inference relations in which a sentence stays (*inferential meaning*) and by investigating how users of the language evaluate the truth-value of the sentence (*referential meaning*). We treat the pragmatic reading of *some* and *most* as their inferential meaning, whereas the logical reading of those quantifiers accounts for the referential meaning. Explaining our results, we attribute the stronger acceptance of inferences from *all* to *most* and from *most* to *some* when compared to acceptance of inferences from *all* to *some* to vagueness of *some* and *most*.

Key words: inference, scalar implicature, default, *most*, *some*

1 Introduction

We discuss our experimental results concerning scalar implicatures of quantifiers such as *some* and *most*. *Scalar implicature* is a result of a pragmatic inference, based on one of the Gricean conversational maxims, i.e. on the *Maxim of Quantity* (Grice, 1989). *Maxim of Quantity* is a commandment to make our contribution to a conversation as informative as required. In other words it means that we should give our interlocutor *maximum* of relevant[1] information. The starting point of our research was an analysis of the so-called *implicational scale* (known also as Horn scale) $\mathtt{Q} =< some, most, all >$, where quantifiers are ordered according to their informativeness. According to the common account (Levinson, 1983) the use of weaker items from the scale *implicates* (i.e. *implies as implicature* or *suggests*) that a sentence with any informationally stronger one is false, since, according to the Griecean *Maxim of Quantity*, a speaker would be required to

[1] Because of the other conversational maxim: *Maxim of Relevance*

make a stronger, more informative utterance if a true one were available. Thus, if one says: (1) "Some girls wear skirts", the quantifier *some* in this sentence implicates that *not all* girls wear skirts. Similarly (2) "Most girls wear jeans" implicates that *not all* girls wore jeans, as both *some* and *most* are weaker on the implicational scale than *all*. What is more, (1) should implicate also that "It is not the case that most girls wear skirts", since *some* is weaker than *most*. We call *not all* the *weak implicature* of *some*, and *not most* the *strong implicature*. On the order hand, the theory assumes that all stronger items from the scale *entail* all the weaker. Thus, *all* entails both *most* and *some*, and *most* entails *some*.

Above we explained the predictions of the pragmatic theory. We must note that all but one of the above-mentioned "entailments" are invalid in the classical meaning of a correct inference. In predicate logic, sentences "Some A's are B" and "All A's are B" are usually interpreted as (accordingly) existentially and universally quantified. Hence the following equivalences hold:

$$All(A, B) \iff \forall x(A(x) \to B(x)) \tag{1}$$

2

$$Some(A, B) \iff \exists x(A(x) \land B(x)) \tag{2}$$

Furthermore, *most* is usually interpreted as logically equivalent to *more than half*:[3]

$$Most(A, B) \iff (|A \cap B| > |A - B|)$$

Given the above definitions, neither "Most A's are B" nor "Some A's are B" logically entail "Not all A's are B". Of course, *some* does not entail *not most* either. This is why implicature is considered to be a result of a pragmatical phenomenon and not a part of semantics.

The considered entailments from the stronger to the weaker items from the scale `Q` are also problematic from the logical point of view. Only the inference from *most* to *some* is logically valid without any restrictions, whereas inferring *some* or *most* from *all* requires additional assumption that the domain is non-empty. This is because, if there are no A's, then "All A's are B" is true but "Most A's are B" or "Some A's are B" are false sentences.

In our experiment, we checked whether the above-mentioned inferences between quantifiers from the scale `Q` are accepted by people. We analyzed implicatures as well as entailments from the stronger items to the weaker. In this way we checked whether this is the pragmatic or the logical reading of quantifiers *some* and *most* that plays the crucial role in reasoning in natural language.

[2] $Q(A, B)$ is a notation used to express the sentence "Q A's are B", where Q is a quantifier from the scale `Q`.

[3] Note that *more than half* cannot be defined in first order logic.

Thus we checked whether our subjects, asked explicitly to evaluate given inferences, would rather respond according to the requirements of logical correctness or according to the pragmatical rules of communication.

2 Experiment

In the experiment we analyzed inferences with positive premises, with quantifiers *all*, *most*, and *some* as well as inference with the negative counterparts of those premises, so with quantifiers with inner negation: *no*, *most not* and *some not*, accordingly. Thus, negative sentences that were used as premises had the following forms: "No A's are B", "Most A's are not B", "Some A's are not B". Tables 1 and 2 summarize the quantifiers used in premises and conclusions of tested inferences. We compared the strong (*most not*)[4] and the weak (*some not*)[5] implicature of *some*. Scalar implicature of *most* was compared in the universal (*not all*) and the particular (*some not*) form. The following were the universal forms of implicature: "Not all A's are B" for the positive premise (*most*) and "It is not the case that no A's are B" for the negative premise (*most not*).

	Premise	*Conclusion*
Positive	*some*	*some not*
	some	*most not*
	most	*some not*
	most	*not all*
Negative	*some not*	*some*
	some not	*most*
	most not	*some*
	most not	*not no*

Table 1. Tested inferences from the weaker items to the negation of the stronger (implicatures)

Note that for all inferences from Table 1 the expected response is "no" if subjects understand quantifiers *some* or *most* in the logical way and "yes" if they understand them in the pragmatical way, so with the implicature. The expected

[4] A reader may observe that "Not most A's are B", which is implicature of "Some A's are B" suggested by the scale `Q` is not logically equivalent to "Most A's are not B", which was actually used in our tests. Given *most* equivalent to *more than half*, *not most* means *half or less than half* and *most not* means *less than half*. Hence "Most A's are not B" is logically stronger but also *simpler* than "Not most A's are B". Because of the minimal difference in the truth-value conditions for both sentences (strict or non-strict inequality) we chose the first sentence as a more linguistically natural candidate for a negative form with *most*. The goal of keeping homogeneity of forms (inner negation) in the experiment was nonetheless important.

[5] *Some not* is logically equivalent to *not all*

	Premise	*Conclusion*
Positive	*all*	*some*
	all	*most*
	most	*some*
Negative	*no*	*some not*
	no	*most not*
	most not	*some not*

Table 2. Tested inferences from the stronger items to the weaker

responses for inferences from Table 2 are opposite: "yes" given the logical reading of quantifiers (with an additional assumption that domains are non-empty in the case when *most* or *some* are entailed by the universal quantifier) and "no" given the pragmatic interpretation.

2.1 Layout

There were four tests, each with sixty inferences to evaluate: forty inferences that were analyzed and twenty fillers. More detailed, in each test there were four different types of tested inferences (each type was repeated ten times with permutated content), and four types of fillers (repeated five times with permutated content).

Subjects were asked whether, on the basis of a given sentence, they can *say* some other sentence.

Example:[6]

You know that

[*wiekszość wyszczyków jest dwukolorowa.*
most wyszczyki is two-colored]
most wyszczyki are two-colored.

Can you say, on the basis of the above sentence, that

[*niektóre wyszczyki sa dwukolorowe.*
some wyszczyki are two-colored.]
some wyszczyki are two-colored?

YES / NO[7]

[6] All the tests were conducted in Polish. Here we give an English translation of a task example, quoting also the original relevant sentence in Polish.

[7] Note that all tests consist of "yes/no" tasks. In the first version of the test, there was a possibility of giving a "don't know" reply. However, based of a pilot research, we observed that subjects never gave this kind of response and chose only "yes" or "no". Therefore, we decided to exclude the "don't know" possibility for the sake of simplifying tests and future statistics.

In the task formulation, there was no reference to such notions as "truth" or "logical correctness". Thus the questions checked behavioral disposition of subjects to accept or not given sentences if some other facts are known to them. In all four tests we used as fillers inferences similar to the analyzed ones but logically incorrect and not justified even in any pragmatical sense. Examples of such incorrect inferences are entailments from *most* to *all*, or from *some* to *no* and similar. It is hard to find any plausible explanation why such inferences (e.g. *most* $\rightarrow$ *no*) would be accepted by people. Thus, there seems to be no model of reasoning in which such an inference would be regarded as correct. The schematic layout of all four tests can be found in the Appendix.

2.2 Choice of Terms

In our experimental tasks we used only so-called *non-sense terms*, thus terms that do not exist in the language, are new-introduced and hence meaningless for our subjects. The reason for this was to reduce the contextual factor in reasoning, so that people would not give answers based on their knowledge about world. It is a known fact that people tend to accept a given conclusion if they think it is true in the light of what they know about the world, or reject it if they think it is false, independently of whether it follows from the premise or not. If a term is meaningless, there is no knowledge about the alleged object it refers to that can be retrieved by subjects from memory. In Appendix we list all terms (ten pairs: a term plus a property) that were chosen for our tests. Note that each pair was used for each syntactic form only once, so that no identical task was repeated. In the case of tasks with only five trials (fillers), we chose five pairs of a term and a property and the remaining five pairs were used for another filler. The tasks in each of four tests were ordered randomly by a computer program.

2.3 Participants

Each of the four tests was solved by an independent group of people. These were students, without logical experience, coming from the following backgrounds: history, English philology, Russian philology, pedagogy, culture studies, criminal investigation techniques, law. Groups were mixed with respect to backgrounds. Table 3 compares group sizes and subjects' age.

GROUP	***1a***	***1b***	***2a***	***2b***
number of subjects	49	48	40	40
age average	22.63	21.16	23.5	22.67
age range	19 – 32	19 – 32	20 – 39	20 – 33

Table 3. Groups

3 Results

3.1 Inferences from the Weaker Items to Negation of the Stronger

We observed that the scalar implicature *some not* (resp. *not all*) was accepted with high frequency for both quantifiers: *most* (over 80%) and *some* (over 90%). Similar result was obtained for the implicature *some* for negative premises: *most not* and *some not*. The significance of this result was confirmed by both χ^2 tests ($p = .000$ in all cases) and analysis of histograms of frequency distribution. Tables 4 and 5 display the results in percentages of accepting a given implicature for a given premise. We can clearly see that the implicature *some not* (resp. *not all*) is accepted for *most* and *some* equally often, however *most not* as entailed by *some* is significantly less frequently accepted by people (ca. 60%). [8]

	some	*some not*
strong	*most not* 57.96%	*most* 53.67%
weak	*some not* 93.33%	*some* 92.3%

Table 4. Scalar implicatures for *some* and *some not*

	most	*most not*
particular	*some not* 80.25%	*some* 81.5%
universal	*not all* 85.25%	*not no* 72.5%

Table 5. Scalar implicatures for *most* and *most not*

A reader may note that the value of frequency of accepting *not (the case that) no* as entailed by *most not* (72.5%) was a bit lower compared to all other

[8] Mann-Whitney tests confirmed the significant difference between the strong (*some not*) and the weak (*most not*) implicatures separately for positive ($Mdn_1 = 10$, $Mdn_2 = 6$, $U = 319.5$, $z = -6.478$, $p < .0001$, effect size:$r = -.658$) and negative ($Mdn_1 = 10$, $Mdn_2 = 5$,$U = 283.5$, $z = -6.628$, $p < .0001$, effect size: $r = -.673$) premises with *some*. Implicatures of *some* and *most* were compared separately for positive and negative premises by the Kruskal-Wallis test (positive premise: $H(3) = 50.629$, $p = .000$ (Asymp. Sig.); negative premise: $H(3) = 48.095$, $p = .000$ (Asymp. Sig.)). Post hoc procedure (Mann-Whitney tests) established significance of the lower effect of the strong (*most not*) implicature of *some* compared to implicatures of *most*: the particular implicature (*some not*) ($U = 458$, $z = -4.393$, $p = .000$) as well as the universal ("not all") ($U = 621.5$, $z = -2.979$, $p = .003$).

values in the table (all over 80%) (Table 5). The analysis proved that this was a significantly lower result compared to the frequency of accepting the logically equivalent form of this implicature (*some*) for the same premise (Mann-Whitney $U = 595$, $z = -2.090$, $p = .036$ (exact 2-tailed), $r = -.233$), as well as compared to the positive counterpart of this entailment, that is *not all* as entailed by *most* (The Wilcoxon signed-rank test's result: $z = -2.602$, $(T = 98)$, $p = .004$ (exact 1-tailed), $r = -.291$). All other relevant comparisons (*most* → *some not* compared with either *most* → *not all* or with *most not* → *some*) were not significant. We cannot conclude that universality of the form of implicature plays any role in people's evaluation of such inferences. However, the lower frequency of accepting *most not* → *not no* inferences can be explained by the complicated syntactic form of the conclusion (*not the case that no*), which uses logical negation twice and a marker of negation in language three times (due to double negation of the form of *no* in Polish).

3.2 Inferences from the Stronger to the Weaker Items

In the case of entailments from stronger items to the weaker, our data revealed a significant disproportion. The striking result was that *some* was accepted as entailed by *all* (resp. *some not* as entailed by *no*) rather rarely – in only ca. 30% of all cases. However, people accepted *most* as entailed by *all* (resp. *most not* as entailed by *no*) in ca. 60% of cases.[9] Hence, our experiment confirms the assumption that weaker items implicate the negation of stronger items, but it does not confirm the assumption that the stronger items imply the weaker ones. What is more, it seems that in the case of items from the opposite poles of the scale (*all* and *some*) this assumption is directly falsified. Particularly interesting is the fact that in the case of the stronger items implying the weaker the frequency of people's acceptance of such inferences seems to reflect distances on the implicational scale, which results in a specific disproportion. Thus, inferring *most* from *all* is as frequent as inferring *some* from *most*, and almost twice as frequent as inferring *some* from *all*. Graphs 1 and 2 illustrate the data for positive and negative premises.

3.3 Analysis of Distribution

An analysis of frequency distribution allows some interesting observations. The frequency of accepting inferences *all* → *most* (resp. *no* → *most not*) displayed a

[9] Mann-Whitney tests proved the significant difference in frequency of accepting entailments from *all* to *some* compared to entailments from *all* to *most* ($Mdn_1 = 2$, $Mdn_2 = 9$, $U = 678.5$, $z = -2.433$, $p = .007$, $r = -.259$) or alternatively to entailments from *most* to *some* ($Mdn_1 = 2$, $Mdn_2 = 6$, $U = 870$, $z = -2.235$, $p = .013$, $r = -.227$). A similar test showed no significant difference between entailments from *all* to *most* and from *most* to *some*. Analogical analysis showed the same for negative counterparts. (Result for *some not* entailed from *no* – compared to *some not* entailed from *most not*: $Mdn_1 = 1$, $Mdn_2 = 6$, $U = 478.5$, $z = -5.075$, $p = .000$, $r = -.515$, and compared to *most not* entailed from *no*: $Mdn_1 = 1$, $Mdn_2 = 8$, $U = 533$, $z = -3.650$, $p = .000$, $r = -.389$.

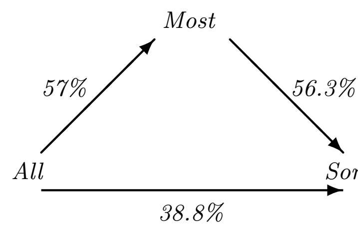

Fig. 1. Graph for positive premises

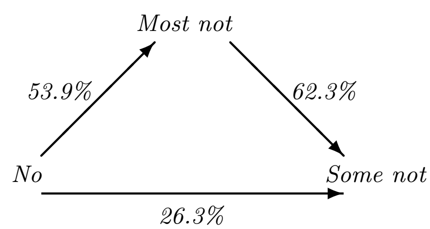

Fig. 2. Graph for negative premises

clearly bimodal distribution with two maxima at values: 0 and 10. In this case it means that subjects tended to reply *coherently* either "no" or "yes" to all questions of this form. This fact and the average result for this category oscillating around $50-60\%$ of "yes" responses suggest that the considered inferences can be dependent on the individual's understanding of quantifiers used in the task. Thus, people tend to have rather clear opinions about such inferences. The histogram for inferences *all* → *some* was again bimodal, however the distribution of the data for *no* → *some not* was clearly positively skewed. The histogram displayed that almost half of people scored 0, so replied coherently "no" to this task.

The data for scalar implicature *some not* (resp. *not all*) for either *most* or *some* displayed a negative skew with a peak at 10. Thus, most subjects scored 10 for this category, so replied "yes" to all questions of this form. This means that, according to our expectations, most subjects recognized such inferences as valid and remained coherent while giving "yes" responses. Indeed, over 75% of subjects replied "yes" in the case of the weak scalar implicature (*some not*) of *some*, remaining coherent in their opinions (so in all possible cases of this task). The same result was obtained for the negative counterpart.

Finally, inferences *most* → *some* (resp. *most not* → *some not*) displayed a tendency to be asymmetric with a slight negative skew (with a peak at 10). The data for the strong implicature of *some* (*most not* entailed from *some* resp. *most* entailed from *some not*) had rather flat distribution, which may be evidence of people's indecisiveness about such inferences.

3.4 Looking for Correlation

Finally, we checked whether subjects' responses to tasks with scalar implicatures were negatively correlated with their responses to corresponding tasks where the weaker items were entailed by the stronger ones. The goal of this analysis was to determine whether it is a subject's individual understanding of quantifiers that plays the role. The bimodal distribution of the data in the case of inferences: *all* → *most*, *no* → *most not* and *all* → *some* could suggest that people's evaluation of such inferences indeed depends on their individual understanding of quantifiers *most* and *some*. This should, however, coincide with the proper treatment of the according implicatures. For example, assume that a subject understands *some* so that it is false in the case of *all*. In such a case, to be consistent, she should reply "yes" to all entailments from *some* to *some not* and "no" to all entailments from *all* to *some*. Then, the claim that *some not* (resp. *not all*) is not implicature of *some* but a part of its meaning would be supported. However, if a subject accepts both inferences, then the phenomenon requires a more complex explantation. Indeed the high acceptance of implicatures in our experiments already precluded the possibility of such correlation, at least in the case of the implicature *some not/not all* for both *some* and *most* which was commonly accepted by almost all subjects independently on what were their opinions about the relevant inferences with *all* as premise and *some* or *most* in the conclusion. The statistical analyses confirmed this prediction: There was no significant (negative) correlation between subjects' scores for accepting scalar implicatures and their scores for accepting the weaker items as entailed by the stronger ones. Thus, there was no correlation between people's frequency of accepting inferences:

- *all* → *some* and *some* → *some not*,
- *no* → *some not* and *some not* → *some*,
- *all* → *most* and *most* → *some not* (resp. *most* → *not all*),
- *no* → *most not* and *most not* → *some* (resp. *most not* → *not no*)

The negative correlation was, admittedly, significant between *some* → *most not* and *most* → *some* inferences (Spearman's rho: $r = -0.373$, $p = .004$, $N = 49$), but it was not significant for the negative counterpart, i.e. *some not* → *most* correlated with *most not* → *some not*.[10]

The lack of correlation between the considered scores excludes the simple explanation that understanding of *some* resp. *most* is dependent on individual differences between subjects who either use these quantifiers with implicatures or without.

[10] The case of correlation is isolated and not conclusive.

4 Discussion

Some of the findings of our research seem quite interesting. First of all the graded nature of the inferences drawn by speakers, i.e. greater acceptance of *all* $\rightarrow$ *most* and *most* $\rightarrow$ *some* inferences in comparison to *all* $\rightarrow$ *some*. Secondly, the very strong acceptance of scalar implicatures, especially compared to the inferences from the stronger to the weaker items, though the latter ones were not commonly rejected. These findings, and especially the disproportional character of subjects' acceptance of inferences from the stronger to the weaker quantifiers may support the vague nature of *most* and *some*. In the discussion of our results we attribute the disproportional result to vagueness of *some* and *most*. We devote also a part of the paper to discussing our findings with regard to two competing accounts of scalar implicatures: the first that implicatures are a result of default inferences (*default theory*) and the second that they are context-driven. Finally, we apply the distinction between the referential and inferential meaning which refers to the way of establishing the meaning in language. Before that, we explain shortly why we do not want to engage the problem of empty domains in the interpretation of our data.

4.1 The Problem of Empty Domains

One could argue that the fact that people so rarely accept *some* as entailed by *all* (resp. *some not* as entailed by *no*) results from incorrectness of such inferences in the case of empty domains. Since we used only non-sense terms, and hence actually all the domains in our tasks were empty in the sense that they lacked reference in the real world, such a claim might seem to have an appealing advantage, namely it would fit nicely the "logical correctness approach" in analyzing human reasoning. Thus, we could say that since *some* has an existential meaning and *all*, as a universal quantifier, lacks this kind of existential load, our subject's logical competence acts precisely in accordance to the logical correctness requirement when they refuse to infer the existential claim from the universal one in the case when the domain is empty. That such an explanation is false can be demonstrated with the below arguments:

- As we have mentioned above, according to the standard definition, both "Most A's are B" and "Some A's are B" are false if there are no A's. Hence, both inferences: *all* $\rightarrow$ *some* and *all* $\rightarrow$ *most* require for validity the additional assumption that the domain is non-empty. However, in our experiment more than half people refused to accept the first inference but they accepted the second one on the level of ca. 60% of all cases – so with the same frequency as the inference *most* $\rightarrow$ *some*, which is valid without any restrictions. If this is the existential load that matters here, then we should also explain why it matters more in the case of *some*, than in the case of *most*.
- Although *some* as entailed by *all* (resp. *some not* as entailed by *no*) was rejected by more than half of subjects, it was still accepted in ca. 30% of all cases (by some subjects even consistently in all possible cases). Thus, the

simple explanation that those inferences were rejected as incorrect fails in the light of the fact that *trivially incorrect* inferences (e.g. $all \rightarrow most..not$) were accepted with much lower frequency (ca. 5%) and rather accidentally in single cases.
– Last but not least, it seems doubtful whether ontological status of terms used in sentences plays any important role in people's inferring sentences with *some* from those with *all*. Another reasoning experiments comparing such inferences showed no significant difference with respect to the sort of terms (empty vs non-empty) used in tasks (Spychalska, 2009). We rather suspect that non-emptiness (fictional or potential) of the domain is somehow presupposed by subjects who then reason as if the "real" ontological status of the entities referred to was not relevant.

4.2 Default- or Context-driven?

In principle, scalar implicatures are what Grice (1989) called *generalized implicatures*, which were assumed to occur systematically although the context might be such that they would not occur. On the other hand, particularized implicatures were assumed to be less systematic and always clearly dependent on context (e.g. "What happened to the chocolate?" "A child has some chocolate on his mouth."). In the literature on scalar implicatures, one can find two theories: that scalar implicatures are default inferences and that they are context-driven. The first one strengthens Grice's suggestion that scalars are context-independent. According to this view scalar implicatures hold by *default*, thus in a normal situation "Some A's are B" will give rise to inference that "Not all A's are B". This inference is linked to the lexical item *some* and is generated more or less automatically and independently on the context, however it may be canceled in special circumstances (Horn, 1984), (Levinson, 2000). Those who reject this assumption (Bott & Noveck, 2004), (Breheny, Katos, & Williams, 2006) claim that scalar implicatures are indeed more similar to particularized implicatures and they are context-driven – they arise in specific context and, thus, are never canceled. For instance *relevance theory* (Wilson & Sperber, 1986) treats such inferences as purely contextual and dependent on a speaker's intention who expects an interlocutor to draw a relevant enough interpretation of an utterance.

In our experiment the high acceptance of implicatures is agreeable with both theories. To advocate the context-driven account, one may say that it is a tacit or *default* context of the experimental situation that makes implicatures to arise. A subject who is given a premise of a form "Some A's are B" may assume that if the experimenter had known anything about all A's, this information would have been given in the test. Thus, one could argue that in fact the implicature is accepted in the experiment based on the default context and not on the defaultness of the implicature itself. However, the data concerning inferences from the stronger items to the weaker are better explained with the assumption that implictures indeed are default-driven. The fact that *some* was in general *rejected* as entailment from *all* may be directly explained by people's understanding of *some* as meaning *some but not all*. The result showing that there is a significant disproportion between frequency with which people would accept *some* as entailed by *all* and *most* as entailed by *all* does not at all seem coherent with the context-driven approach to implicatures. Indeed, if implicatures arise because of a specific context of utterance, the traditional semantics of quantifiers *some* and *most* is preserved and hence the inferences *all* $\rightarrow$ *some*, *all* $\rightarrow$ *most* or *most* $\rightarrow$ *some* are still valid (given the already mentioned restriction about non-emptiness of the domain). There is nothing in the context of this experimental situation: explicit or default that should cancel validity of such inferences. Actually, the advantage of the context-driven solution was to preserve traditional semantics, and hence the above-mentioned inferences, while allowing also implicatures, so inferences that are incorrect from the logical point of view, to arise in specific contexts based on pragmatical factors. If we, however, lose inferences that base on traditional semantics we may ask if this semantics is indeed the right one. Let us observe that – tacit or not – the context was exactly the same in the case of all three kinds of inferences. Why then people were more willing to accept *most* as entailed by *all* or *some* as entailed by *all* than *some* as entailed by *all*? The context-driven theory does not explain this result.

Does the default theory deal with the experimental outcome better? If *not all* is a default inference from *some* and *most*, then it is clear why such entailments are commonly accepted. The difficult part concerns explanation of the disproportion figured in Graphs 1 and 2. Saying that scalar implicatures arise by default and are connected to the lexical items *some* and *most*, and not to the context, may be interpreted in this way that *not all* is already a part of some default meaning of those quantifiers. Then, however, both *some* and *most* should be, for the sake of consistency, equally rejected as entailments from *all*. However, as we have observed, while the first one is indeed rejected by over 60% of our respondents, the latter is, on the contrary, accepted by 60%. Thus, there has to be some reason why cancelation of implicature in the case of *most* is easier than in the case of *some*. Otherwise we need a better theory.

Below, we propose a kind of modification of the default-driven approach which takes into account vagueness of quantifiers and differentiates between so-called *inferential* and *referential meaning*. Before we proceed in describing our proposal, we would like to refer to some experimental data that are used to reject the default theory of implicatures.

4.3 Are Default Inferences Automatic?

It is frequently argued that if the default theory was true, than the default reading of a quantifier expression (i.e. *some* as *some but not all*) would require shorter processing time than the reading in which the implicature has to be canceled. Empirical data are reported to support this claim and reject the default theory, e.g. Bott & Noveck (2004). Authors conducted four experiments employing a sentence verification paradigm with a measure of a reaction time. They focused on underinformative sentences such as “Some monkeys are mammals” which are false given implicature *not all* is a part of the meaning of *some* (*pragmatic interpretation*) and true given *some* means *some and possibly all* (*logical* *interpretation*). They report on results showing that the pragmatic reading of this sentence, even if it is chosen by people in the first instance, requires longer reaction time and hence more processing effort. In this way, the authors argue that the pragmatic reading of *some* is not automatic or default, but it is rather the logical interpretation, which is default and comes effortless. It is concluded that if implicatures require additional processing effort, they cannot be default inferences.

The experiments by Bott & Noveck show in an elegant way that the pragmatical reading of *some* is processed longer than the logical reading. What we would like to discuss, and what comes along with those experimental data, is the ambiguous, in our opinion, sense in which implicatures are considered as default or not default. What does it mean that implicatures are default? Originally, so in Levinson's and Horn's works, it is linked to context-independence and is explained in terms of automatic inference. Such theorists as Bott & Noveck who argue against the default status of implicatures emphasize that "default" means "automatic", where "automatic" is however not a well clarified notion but since it is connected to expected shorter processing time, it may be understood in the sense of automaticity of some cognitive processes. Thus, the argument structure is the following: if implicatures are default inferences (since they are context-independent), this means that they are automatic inferences, and "automatic" means "easiest to process". Then, it is observed that the pragmatic reading of *some* takes longer time that logical reading, and hence it is concluded that the scalar implicature cannot be automatic or default.

We argue here that the default character of implicatures does not necessarily mean that they are automatic, especially if they are context-independent. In our proposal, a *default meaning* refers to the *preferred meaning* for an ambiguous word, and a *default inference* to the *most common inference* that is entailed (from a given premise) by language users. However, "most common" does not necessarily mean "automatic" or "easiest to process". The reason why some inferences are default in the above sense may be that they make communication more efficient because statistically they lead us to the truth, even if they require more processing time.

4.4 Inferential and Referential Meaning

It is worth pointing out that the default character of some inferences, e.g. *some* $\rightarrow$ *not all*, is frequently identified with the default meaning of the lexical items – in this case with the meaning of *some*: *some but not all.* However, drawing a conclusion and understanding language are probably different cognitive processes. Moreover, the active inference generation may differ from the passive evaluation of validity of inferences and also from the passive evaluation of the truth-value of sentences. We appeal here to the distinction between *inferential* and *referential meaning* introduced in (Mostowski & Wojtyniak, 2004) as a plausible solution to the problem of computational complexity of some natural language sentences, like branching sentences (e.g. "Hintikka sentences"). While referential meaning is established by investigating how users of the language evaluate the truth-value of a sentence in various situations, inferential meaning is assigned by inference relations in which this sentence stays. Mostowski & Wojtyniak (2004) propose that referential and inferential meanings of some natural language expressions may differ or that some language expressions may get its meaning determined mainly through inference relations.

We suggest that implicatures can be also analyzed in terms of inferential and referential meaning. In our proposal we assume that:

- *Not all* is inferred by default from both *some* and *most*.
- Inferential and referential meanings of *some* and *most* differ; the inferential meaning is established through inference relations that are accepted by language users. Note that we extend the notion of "inference" here so that it does not only refer to logical inferences but also to pragmatical ones, i.e. scalar implicatures.
- *some* and *most* are vague quantifiers; while *most* is used mainly for proportions around 80%, (inferential) meaning of *some* implicitly conveys *not many*.

Given *not all* is the default inference from *some* and *most*, then the *inferential meaning* of those quantifiers will be exactly the pragmatical reading. Of course this process (of recognizing the inferential meaning) cannot be effortless, since people have to go through the inference, for example to decide about the truth-value of underinformative sentences like "Some elephants are mammals" – as in Noveck's and Bott's experiment. The fact, however, that rejecting such sentence, based on the pragmatical interpretation of *some*, is longer than accepting it given the logical reading of *some*, does not mean that the first reading cannot be the default one. In Noveck's and Bott's experiment ca. 60% of people responded that "Some elephants are mammals" is false, whereas only 40% admitted it is true (Bott & Noveck, 2004). Even if those 60% took longer time to process the sentence, it does not mean that their reply was somehow context-dependent. It was not automatic, but it was default in this sense that it was preferred. The reason why people choose the inferential meaning here, is that such choice is probably optimal from the point of view of effective communication. Language users predict what kind of inference their interlocutors would make from a given sentence and they predict what their interlocutors predict about their own predictions in this matter.

A reader may observe that, at least in the case of the considered quantifiers *some* and *most*, referential meaning can be identified with logical reading and inferential meaning with pragmatical. The goal of using the new notions is, first of all, that the inferential-referential distinction is more universal and may be applied also in cases when no implicature is involved; secondly, it is to switch the point of view – instead of ambiguous semantics we have rather two different *ways* of establishing meaning: via verifying the truth-value of sentences and via recognizing its inference dependencies. The resulting meanings obtained by each of those two methods do not have to be identical, but they will play role in different contexts and different communication situation. As it turns out the inferential meaning will be much more important in a common, daily language.

Finally, we may refer to the disproportion figured in the triangle graphs: 1 and 2. Given the implicature *not all* is a default inference from *some* and *most*, these quantifiers are not used if application of the quantifier *all* also results in a true sentence. The reason is that the speaker will not want her interlocutor to infer an unwanted *not all* conclusion. As mentioned above, she bases on her predictions about what her interlocutor will probably infer from her claim. Similarly, a person who is given a sentence with one of those two quantifiers usually interprets this sentence as conveying also the considered implicature. This, however, bases on a meta-prediction, i.e. I predict that my interlocutor predicts that I would infer *not all*, so to hinder me from doing it, she should have told me that *all* if she had known that this is the case, or she should have warned me that she did not have full information. In our experiment, the situation seems to change when people are asked to evaluate sentences with *some* or *most* as true or not in the light of universal premises. Since inferences: $all \rightarrow most$ and $all \rightarrow some$ become valid assuming the referential meaning, this fact can suppress using the inferential meaning. Thus, those inferences can be treated in two ways, namely the conclusion may be rejected based on the inferential meaning of *some* or *most*, or the meaning may be changed to the referential *at least some (most)*, and thus *some* or *most* will be understood without the implicature. Since *some* is not anymore in the premise, but in the conclusion and the premise clearly states that *all* holds, there is no need to predict anything about predictions about inferences that may be made by our interlocutors. Everyone knows that *all* is the case. We are suppressed from drawing *not all* from *some*. The essential point is what cognitive mechanism lies behind this (second) interpretation, and especially why it occurs more often in the case of *most* than in the case of *some*. Here we appeal to vagueness of *some* and *most*.

4.5 Vagueness

It is argued in the linguistic literature, opposite to the traditional and logical approach, that the way *most* is used in natural language differs much from the simple *more than half*. Whilst *more than half* is acceptable for proportions close to 50%, *most* is dispreferred in such cases (Ariel, 2004) (Ariel, 2005). Finally, *most*, opposite to *more than half*, gets usually a generic interpretation, that is it refers to the whole kind or typical representative. Thus, the use of this quantifier is not only vague in the sense of cardinality (*most* will not be typically used for 51% but will be for 80%), but is also dependent on the context (e.g. is 80% or 70% a typical representative for a given domain?).

While vagueness of *most* has been already recognized, there is not so much data about plausible vagueness of *some*, though one can find suggestions in the literature that *some* entails *a few* (Channel, 1994). There exist also data which seem to suggest that *some* is dispreferred for describing situations in which one refers to more than half objects in a given domain. Spychalska (2009) presents results of a sentence-generation experiment with the use of picture models (white and black dots). Subjects were asked to write down all the sentences of a given form ("*Q* dots are/are not black") with quantifiers *some*, *most*, *all*, positive or negative, which they find true about the pictures. It turned out the the more dots were black in the model, respectively the higher percentage of dots were black in the model, the less people wrote that "Some dots are black". The (negative) correlation between the number of black dots and the number of people who gave a response with *some* was significant and very strong (Spearman's rho $r = -.598$; $p = .007$, $N = 16$ [11]), and even stronger if the variable was the percentage of black dots (Spearman's rho $r = -.736$; $p = .001$,$N = 16$). Analogously there was a strong (positive) correlation between the number of back dots and the number of responses with *some not* (Spearman's rho $r = .820$; $p = .000$, $N = 16$), similarly for the percentage of black dots in the domain (Spearman's rho $r = .849$; $p = .000$, $N = 16$). The above result can suggest, although it is not sufficient to be used as evidence, that there is indeed a kind of vagueness in the meaning of *some*, which is a quantifier less willingly used when referring to bigger samples of objects (or bigger subsets of fixed domains). If *some* by default implies *not many* (or *not many with respect to a domain given*), it may explain people's indecisiveness about *most not* as an entailment from *some* (a flat distribution of data and the result oscillating around 60%). Thus, this kind of vagueness of *some* may explain both: the result for the strong scalar implicature of *some* (inferences *some* → *most not*) as well as for the inferences *most* → *some* (again a flat frequency distribution and the result on the level of ca. $50 - 60\%$). Note that we do not claim that the referential meaning of *some* or even *most* is vague. Both *some* and *most* may have perfectly sharp meaning while they are taken in their referential sense. Vagueness will, however, display in the active language use, and in what people will infer from sentences with *some* and *most*. This property might be investigated experimentally. For instance in experiments using a verification procedure both *some* and *most* should be accepted for all borderline proportions, whereas both: production of sentences describing given models as well as generating inferences with *some* and *most* should display a vague attribution of given proportions as typical or not typical for each these quantifiers.

To summarize: The clear bimodal distribution of data for *all* → *some* as well as *all* → *most* inferences suggests that people were consistently choosing one of the two readings. Still, the choice of the pragmatic (i.e. inferential) reading in the case of *some* was more frequent than in the case of *most*. The fact that *most* is preferred for proportions close to 80% may explain why more people were willing to infer *most* from *all* than *some* from *all*. Even if we tend to infer *not all* from *most*, we also tend to use *most* for generic claims. Thus, *most* may be sometimes interpreted as the whole domain. However, in the case of *some* we not only tend to infer *not all*, but also, although to a lesser extent, *not most*. This creates a gap between the referential and inferential meaning – grater in the case of *some*, than in the case of *most*. As a result, the logical reading of *some* may be much stronger suppressed from occurrence in reasoning than is the logical reading of *most*.

[11] All the p-values are for one-tailed correlations

5 Conclusion

In the paper we have presented and discussed the results of our reasoning experiments concerning scalar implicatures of quantifiers *some* and *most*. In our proposal scalar implicatures are context-independent and default in this sense that the pragmatic interpretation of the lexical item with which they are connected is preferred in the communication. We treat hence the pragmatic reading of *some* and *most* as their inferential meaning, thus meaning established through the inference relations, which are recognized by language users for given lexical items. The inferential meaning is contrasted to the referential meaning, which is the logical reading of those quantifiers. We use the inferential-referential distinction to emphasize that the two meanings are established in different ways. We observe that the inferential meaning of both considered quantifiers is vague. The vague use of *some* which is a quantifier dispreferred for proportions over 50%, and vagueness of *most* which is preferred for those close to 80% as well as given often generic interpretation are the reasons why the pragmatic reading of *some* is more frequently chosen while evaluation validity of *all* → *some* inferences, than is the pragmatic reading of *most* in the case of inferences *all* → *most*. Vagueness of *some* explains also the result concerning people's indecisiveness about the implicature *most not* for this quantifier.

Acknowledgments. I would like to thank to Michiel van Lambalgen for his comments and advises, to Jarmo Kontinen for very useful remarks, to all the people who helped me with the experiments, and to the anonymous reviewers.

Bibliography

Ariel, M. (2004). Most., *Language*, 80, 658706.

Ariel, M. (2005). A "just that" lexical meaning for *most*. In: K. von Heusinger & K. Turner (eds.) *Where semantics meets pragmatics* (pp. 49-91). Amsterdam: Elsevier.

Bott, L., Noveck, I.A. (2004). Some utterances are underinformative: The onset and time course of scalar implicatures., *Journal of Memory and Language*, 51(3), 437-457.

Breheny, R., Katos, N. & Williams, J. (2006). Are generalized implicatures generated by default? An on-line investigation into the role of context in generating pragamtic inferences. *Cognition*, 100 (3). pp. 434-463.

Channel, J.(1994). *Vague Langauge*. Oxford: Oxford University Press.

Grice, H.P. (1989). *Studies in the Ways of Words.*, Cambridge, MA: Harvard University Press.

Horn, L.R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicature. In: Schiffrin, D. (ed.), *Meaning, Form, and Use in Context: Linguistic Applications.*, GURT84, 11-42, Washington; Georgetown University Press.

Levinson, S. C. (1983). *Pragmatics.*, Cambridge, England: Cambridge University.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature.*, MIT Press: Cambridge, Mass.

Mostowski, M. & Wojtyniak, D. (2004). Computational complexity of some natural language construction., *Annals of Pure and Applied Logic*, 127, p.219-227.

Spychalska, M. (2009). *Scalar Implicatures and Existential Import: Experimental Study on Quantifiers in Natural Language.*, ILLC Publications Series, MoL-2009-10. (http://www.illc.uva.nl/Publications/ResearchReports/MoL-2009-10.text.pdf)

Wilson, D. & Sperber, D. (1986). Inference and Implicature. In. Travis, C. (ed.), *Meaning and Interpretation*, pp. 45-76. Oxford: Basil Blackwell.

Appendix

Tests schematic layout:

[**Test *1a***]

- Tested inferences:
  - *most* → *some* (10)[12]; *most not* → *some not* (10)
  - *some* → *most not* (10); *some not* → *most* (10).
- Fillers:
  - *most* → *most not*(5); *most not* → *most* (5)
  - *some* → *some not* (5); *some not* → *some* (5)

[**Test *1b***]

- Tested Inferences:
  - *all* → *some* (10); *no* → *some not* (10)
  - *some* → *some not* (10); *some not* → *some* (10)
- Fillers:
  - *some* → *all* (5); *some not* → *no* (5)
  - *all* → *no* (5); *no* → *all* (5)

[**Test *2a***]

- Tested Inferences:
  - *most* → *some not* (10); *most not* → *some* (10)
  - *all* → *most* (10); *no* → *most not* (10)
- Fillers:
  - *most* → *all* (5);*most not* → *no* (5)
  - *all* → *most not* (5); *no* → *most* (5)

[**Test *2b***]

- Tested Inferences:
  - *most* → *not all* (10); *most not* → *not no* (10)
  - *all* → *most* (10); *no* → *most not* (10)
- Fillers
  - *most* → *all* (5); *most not* → *no* (5)
  - *all* → *most not* (5); *no* → *most* (5)

List of terms:
mermogliny + pink
buzaki + green
mroczniaki + bring bad luck
grombliny + have claws
mgłowce + have a cap
zarkotki + have long ears
trakloki + intelligent
wyszczyki + two-colored
klawuchy + have red tails
leprokraki + like cheese

[12] In brackets: the number of examples