



Verification Strategies for Two Majority Quantifiers in Polish *

Barbara Tomaszewicz
University of Southern California
barbara.tomaszewicz@usc.edu

Abstract. We provide experimental evidence suggesting that the logical structure of linguistic expressions can directly influence aspects of non-linguistic cognition. Specifically, we show that quantifier semantics provides a set of instructions to visual verification processes. Each of the two Polish majority quantifiers *większość* and *najwięcej* biases a distinct verification strategy that is used as a default under time pressure. Each canonical verification strategy overrides other alternative strategies for truth verification as proposed in Lidz *et al.* (2009).

1 Introduction

Lidz *et al.* (2009) propose that the lexical semantics of natural language quantifiers is transparently associated with canonical procedures for the verification of the truth/falsity of sentences in which they appear. In particular, Lidz *et al.* (2009) and Pietroski *et al.* (2008) provide experimental evidence that when processing the proportional quantifier *most* in the context of a visually presented scene, English speakers are biased towards using a certain verification strategy rather than an alternative. This is taken to be evidence in favor of a particular semantic representation of *most*, which in turn provides a direct set of instructions to the visual system that can override other cognitively available verification strategies.

We provide further experimental evidence that quantifier semantics is transparently associated with a canonical verification strategy. We tested the processing of two majority quantifiers in Polish in a task similar to that of Lidz *et al.* The proportional *większość* has the semantics of English *most*,

* I would especially like to thank Roumi Pancheva and Elsi Kaiser, as well as Barry Schein and Toby Mintz, for their invaluable comments and/or help with the experiments. Thanks are also due to Hagit Borer, Jeff Lidz, Victor Ferreira, Jon Gajewski, Robin Clark, Tom Buscher, Katy McKinney-Bock, Mary Byram, Ed Holsinger, Krzysztof Migdalski, Dorota Klimek-Jankowska, Grzegorz Jakubiszyn, Ewa Tomaszewicz, Marcin Suszczyński and the students of IFA at Wrocław University and WSF Wrocław. My work was partially supported by an *Advancing Scholarship in the Humanities Grant* from USC to R. Pancheva.

while the relative *najwięcej* has the meaning of ‘the largest subset’. We obtained three notable results. First, the proportional *większość* is associated with the same pattern of accuracy as English *most*, directly replicating the findings of Lidz et al. for Polish. Second, the relative *najwięcej* is verified by a distinct strategy directly related to its semantics. Third and most important, each strategy is canonically followed during the processing of the respective quantifier. In principle, given that both strategies are easily available to speakers of Polish, when participants judge a scene, where either would yield the correct truth value, they could switch to the more efficient strategy. This is, however, not what happens: participants always perform in accordance with the strategy associated with the quantifier in the stimulus sentence.

Our findings illustrate that two closely related quantifiers can be associated with distinct verification procedures, in line with their lexical semantics, and that each procedure is used consistently as an instruction to the visual system. The conclusion is that the logical structure of linguistic expressions can directly influence aspects of non-linguistic cognition.

2 Background Research Question

Lidz et al. (2009) advance a novel hypothesis that there is more to meaning than just empirical *adequacy* and *compositionality*. There can be several truth-conditionally equivalent compositional specifications of a linguistic expression, but not all of them form equally “good psychological hypotheses” about how the derived truth-conditions are verified.¹

The proportional quantifier *most* can be specified in at least three truth-conditionally equivalent ways, as shown in (1). Pietroski et al. (2008), Lidz et al. (2009), and Hackl (2009) devised experiments to look “beyond” the truth conditions of (1) to see how the meaning of a sentence containing *most* constrains the way people verify it against a visual scene.

- (1) Most of the dots are yellow.
- (a) $|\text{Dot}(x) \ \& \ \text{Yellow}(x)| > 1/2 |\text{Dot}(x)|$
 - (b) $|\text{Dot}(x) \ \& \ \text{Yellow}(x)| > |\text{Dot}(x) \ \& \ \sim \text{Yellow}(x)|$
 - (c) $\text{OneToOnePlus}(\{\text{Dot}(x) \ \& \ \text{Yellow}(x)\}, \{\text{Dot}(x) \ \& \ \sim \text{Yellow}(x)\})$

The semantic specifications in (a) and (b) both provide instructions to the visual cognition system to estimate the cardinality of the set of yellow dots

¹ Note that judging a sentence to be true/false in a given context involves: (i) compositionally determining what the truth conditions are; and (ii) determining whether these conditions obtain in the context. This means that verification procedures can in principle be independent of the algorithms that produce truth conditions, as discussed in Pietroski et al. (2008).

and to compare it with the cardinality of another set. They differ in what that other set is. (1a) requires that the cardinality of the total set of dots be obtained (and its half calculated). In effect, it calls for an algorithm equivalent to that of *More than half of the dots are yellow*. (1b), on the other hand, is linked to an algorithm for verification that requires an estimate of the cardinality of the non-yellow set (which may employ an estimate of the total but does not need to). The alternative in (1c) does not require an estimate of cardinalities or comparison, but relies on matching the yellow dots with the non-yellow dots. The strategies in (1a-c) are semantically equivalent, but not all of them turn out to be psychologically viable options for the verification of the truth value of sentences containing *most* against visual stimuli of arrays of dots.

Hackl (2009) used a self-paced counting paradigm with rows of dots in two colors to establish that *most* and *more than half* are processed differently. His results exclude (1a) as a representation of the meaning of *most* and, consequently, as a verification strategy associated with *most*, at least as far as explicit counting is involved.² Pietroski et al. (2008) further tested the two alternative options in (1b-c) and found that even when the arrangement of dots favored verification by the one-to-one correspondence relation (dots were arranged in pairs, with some yellow dots unpaired with the dots in the other color), the response accuracy patterns did not differ from the condition where the dots were scattered on the screen. No change in accuracy patterns across conditions indicates that (1c) was never used to verify (1).

(1b) can be straightforwardly used to verify (1) when the displayed dots are in two colors only, e.g., yellow and blue, as they were in the experiments of Hackl (2009) and Pietroski et al. (2008). The cardinality of the target yellow set can simply be compared to the cardinality of the blue set, i.e. the non-yellow set in (1b). When the non-yellow set contains dots of multiple colors, obtaining its cardinality requires further computation. Lidz et al. (2009) used multiple colors in their experiment to test whether this computation is based on the components directly expressed in the meaning of (1). Lidz et al. propose that the second argument of the “>” relation in (1b) can be transparently computed by subtraction as stated in (2a) below. Otherwise, the set of all non-yellow dots has to be selected as specified in (2b).

² However, the assessment of cardinality does not always require explicit counting but can be obtained by the Approximate Number System (ANS) (Dehaene 1997, Feigenson et al. 2004). Halberda et al. (2008) have shown that children who have not yet learned to count are perfectly able to understand sentences containing *most*.

- (2) Two alternative computations of
 $|\text{Dot}(x) \ \& \ \text{Yellow}(x)| > |\text{Dot}(x) \ \& \ \sim \ \text{Yellow}(x)|$
 (a) Subtraction strategy:
 $|\text{Dot}(x) \ \& \ \text{Yellow}(x)| > |\text{Dot}(x)| - |\text{Dot}(x) \ \& \ \text{Yellow}(x)|$
 (b) Selection strategy:
 $|\text{Dot}(x) \ \& \ \text{Yellow}(x)| >$
 $|\{\text{Dot}(x) \ \& \ \text{Red}(x)\} \cup \{\text{Dot}(x) \ \& \ \text{Blue}(x)\} \cup \{\text{Dot}(x) \ \& \ \text{Green}(x)\} \cup \dots|$

Lidz et al. (2009) point out that the Selection procedure in (2b) is not plausible for psychophysical reasons. A heterogeneous set of non-yellow, multi-colored dots that are scattered among yellow dots is not automatically selectable as its specification involves a negation of an early visual feature, the color yellow (Wolfe 1998). The Subtraction procedure in (2a), on the other hand, is based on the psychological evidence from Halberda et al. (2006) that multiple color sets can be enumerated in parallel, but crucially, this is possible only for the total set of dots and two color subsets (i.e. total, target and one color distractor sets), but no more.

Given this psychophysical evidence, Lidz et al. (2009) hypothesize that *most* is verified using the Subtraction strategy in (2a), at least in the general case. The strategy involves the following steps: (i) selecting the superset of all dots and estimating its cardinality; (ii) selecting the set of yellow dots and estimating its cardinality; (iii) subtracting the cardinality of the yellow set from that of the superset to obtain an estimate of the cardinality of the set of non-yellow dots; and (iv) comparing the cardinalities of the sets of yellow and non-yellow dots. Since the selection of the superset and one color subset is done automatically, the Subtraction strategy should always be available, independently of how many color sets there are on the screen. However, on screens with dots in only two colors, Selection becomes a viable strategy as well. Given the findings of Halberda et al., both the yellow set of dots and the distractor color set of e.g., blue dots, are automatically selected and their cardinalities can be directly compared. Moreover, in this special case Subtraction involves more steps than Selection and thus may turn out to be dispreferred (see (3)).

- (3) The steps in the computation of Subtraction vs. Selection

Subtraction (irrespective of no. of colors)	Selection (two colors)
1. Estimate the total .	1. Estimate the target set.
2. Estimate the target set.	2. Estimate the distractor set.
3. Subtract the target set from the total.	3. Compare with the target set.
4. Compare the difference with the target set.	

Lidz et al. (2009) presented participants with the sentence *Most of the dots are blue*, and they had to judge it true or false against a visual display of dots in 2 to 5 colors in varying ratios of blue to non-blue dots. The array of dots appeared on the computer screen for 150ms.³ Lidz et al. predicted that if participants use the Selection strategy they should be successful when there are only two colors on the screen. With higher numbers of colors, their performance should rapidly decline, given that they would need to determine the cardinality of each subset of non-blue dots (e.g. red, green, etc.) and subsequently sum the results. The Subtraction hypothesis, on the other hand, predicts no difference in accuracy between screens with dots in two colors and those with 3-5 colors, because the cardinality of the non-blue set is obtained solely on the basis of the cardinality of the total and the blue sets.

The results of Lidz et al.'s experiment support the Subtraction hypothesis because the participants' performance did not differ in accuracy as a function of the number of colors in the display, but only as a function of the ratio (in adherence to Weber's law). Crucially, on screens with just two colors, the alternative Selection strategy is in principle available to the visual system, and it would even be computationally less costly and more accurate (cf. (3)). Yet, even here Subtraction was used, since the accuracy was not higher on the two color screens. Thus, Lidz et al. conclude that Subtraction is the default procedure for verifying *most* under time pressure. On the basis of this finding they formulate the Interface Transparency Thesis (4):

- (4) "A declarative sentence is semantically associated with a canonical procedure for the verification of its truth value that is biased towards those algorithms that directly compute the relations expressed in the meaning." (Lidz et al. 2009: 2)

3 Polish *Most1* and *Most2* Majority Quantifiers

We address the question why Subtraction, as in (2a), is the verification strategy for (1). One reason could be that under time pressure, the Selection strategy (2b) is only possible when there are two color sets, given the findings of Halberda et al. (2006). The Subtraction strategy (2a) is usable under time pressure independently of the number of distractor color sets, and because of

³ Under time pressure, counting is impossible, as typical rates of overt and covert counting have been reported to be about 240ms per item (Whalen et al. 1999). The ANS is not subject to such speed limitations. It generates an approximate representation of the number of items in a set in adherence to Weber's law: the discriminability of two quantities is a function of their ratio. Numbers can thus be represented as 'noisy magnitudes' for the purposes of basic arithmetic operations like addition and subtraction (Wiese 2003).

this universality it can be adopted as the canonical verification strategy for *most*. On this view, Selection in (2b) is excluded not by the semantics of *most* but by the properties of visual cognition. We show that the Selection procedure is possible under time pressure, with more than two colors if performed step-wise, yet it is used only when directly specified in the meaning.

We extend the predictions of the Interface Transparency Thesis to crosslinguistic data, providing evidence that the Polish counterpart of the English quantifier *most* also comes with a verification strategy defined by Subtraction as in (2a). The result is further – and directly – supported by a control condition with a closely related quantifier, which unambiguously requires a Step-wise Selection strategy defined below.

(5) Step-wise Selection strategy:

- $$\begin{aligned} |\text{Dot}(x) \ \& \ \text{Yellow}(x)| &> |\text{Dot}(x) \ \& \ \text{Red}(x)|, \ \& \\ |\text{Dot}(x) \ \& \ \text{Yellow}(x)| &> |\text{Dot}(x) \ \& \ \text{Blue}(x)|, \ \& \\ |\text{Dot}(x) \ \& \ \text{Yellow}(x)| &> |\text{Dot}(x) \ \& \ \text{Green}(x)|, \ \& \ \dots \end{aligned}$$

This control condition in a within-subjects design, where the same group of participants is tested on both items, provides additional evidence for the Interface Transparency Thesis. Comprehenders appear to be biased towards the use of one particular verification strategy that is associated with a given lexical item. They continue to use it even when an alternative strategy, biased by a closely related item, is cognitively available and could even be less computationally costly.

Polish has two majority quantifiers: *większość* (from now on *Most1*) is a counterpart of English *most*, while *najwięcej* (*Most2*), has the meaning of “the largest subset”. *Most2* is true when the cardinality of the target set is greater than the cardinality of each of the distractor sets separately; therefore its interpretation necessarily involves multiple selection and comparison with each distractor set. The two quantifiers are closely related morphologically.

(6) The morphology of *Most1* and *Most2*

(a) *Most1*: *większość*, ‘majority’

<i>więk-</i>	<i>-sz-</i>	<i>- ość</i>
‘many/great’	adjectival comparative ‘-er/more’	nominalizer

(b) *Most2*: *najwięcej*, ‘largest subset / the most’

<i>naj-</i>	<i>-więc-</i>
adverbial superlative ‘-est/most’	‘many/great’
<i>-ej</i>	
adverbial comparat. ‘-er/more’	

(c) *wiel-e / więc-ej / naj-więc-ej;* *wielk-i / więk-szy/ naj-więk-szy*
 many more most great greater greatest

3.1 Materials and Methods

We conducted an on-line visual verification task, asking twenty native speakers of Polish to evaluate the truth of (7) and (8) against 200ms displays of arrays of colored dots, manipulating (i) the ratio between the target color set and the (largest) distractor sets and (ii) the number of distractor color sets.

(7) *Większość kropek jest żółta.*
Most1 dots is yellow
 ‘Most dots are yellow.’

(8) *Najwięcej jest kropek żółtych.*
Most2 is dots yellow
 ‘Yellow dots are the largest subset.’

Each participant judged 360 displays presented in 2 blocks (180 for each quantifier, half requiring a ‘yes’ and half a ‘no’ response). Participants saw the test sentence for 7s, and after each stimulus was flashed for 200ms, they had 3.8s to respond ‘yes’ or ‘no’ by a button press. Yellow dots were present on every display, together with 1 (e.g., Fig. 2), 2 or 3 (Fig. 1) other distractor color sets. Ratios of yellow and non-yellow dots were 1:2 (Fig. 2), 2:3 (Fig. 1) or 5:6.⁴

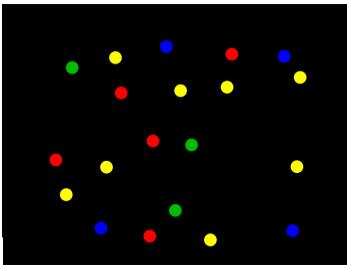


Fig. 1

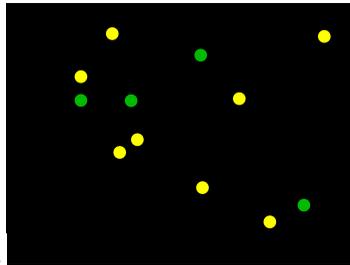


Fig. 2

⁴ On the true screens for *Most1* the cardinality of the yellow set was more than half of the total. For *Most2* the yellow set never exceeded the half but was bigger than the largest competitor color set. False screens were constructed analogously. Thus, *Most1* and *Most2* had identical screens only on the two color set condition. If with *Most2* the largest subset was bigger than the half of the total, we would have no way of checking if the participants switched between strategies.

3.2 Predictions

By the Interface Transparency Thesis (4) each quantifier should come with its own verification strategy, which should be used even when the alternative would yield the correct truth value. There is a body of psychological evidence pointing to the fact that the selection of a target by the visual system is affected by the strategy adopted either consciously or unconsciously (Trick 2008). From this perspective, an empirical finding that *Most1* and *Most2* induce selective attention differently would underscore the importance of fine-grained compositional semantics in real-time sentence processing.

Given that Polish *Most1* has exactly the same meaning as English *most*, we predicted that the Polish sentence in (7) should result in the same pattern of accuracy.

The semantics of *Most2* involves Stepwise Selection of each color set and comparison between the target set and each distractor set as defined in (5). Therefore, we expected to find a significant effect of distractor in addition to a significant effect of ratio. The semantic specification of *Most2* suggests that selective attention should discriminate more than two target color sets, but if this is not possible under time pressure, the performance on the screens with more than two colors will greatly decline as hypothesized by Lidz et al. (2009).

Direct comparison of *Most1* and *Most2* on the screens with dots in two colors can have two predicted outcomes. Since both strategies are used by the speakers of Polish, on two color screens participants could use whichever strategy is computationally less costly and more accurate under time pressure. The computation by Selection requires fewer steps than Subtraction when there are dots in two colors only, as shown in (9).

(9) Subtraction procedure and Stepwise Selection procedure

(a) Subtraction (irrespective of no. of colors)	Stepwise Selection	
	(b) multiple colors	(c) two colors
1. Estimate the total .	1. Estimate the target set.	1. Estimate the target set.
2. Estimate the target set.	2. Estimate 1st distractor set.	2. Estimate 1st distractor set.
3. Subtract the target set from the total.	3. Compare with the target set.	3. Compare with the target set.
4. Compare the difference with the target set.	4. Estimate 2nd distractor set.	
	5. Compare with the target set.	
	6. Estimate 3rd distractor set.	
	7. Compare with the target set.	
	8. ...	

If the semantic representation guides verification, then with *Most2* the non-yellow set should be selected directly and the accuracy should be greater than with *Most1* where the non-yellow set is computed through Subtraction. Alternatively, if the accuracy is the same with both quantifiers on the same screens, it would mean that Subtraction is not used globally for *Most1* and participants are able to switch to the Selection strategy in favorable circumstances. The crucial findings for English *most* were that the OneToOnePlus pairing and the Selection strategy were *never* used. In Polish, however, Stepwise Selection is used for *Most2*, which makes it possible that participants can switch between Subtraction and Selection for the two color screens where the two quantifiers lead to the same truth value. Such a finding would argue against the Interface Transparency Thesis.

If participants are overall successful with *Most2* just as with *Most1* but their performance differs on two color screens, it would suggest that it is not the case that Selection is cognitively ‘harder’, but that it depends on how selective attention is induced by the specifics of the linguistic input. Such a result would also suggest that canonical verification strategy is directly computable from the relations specified in the semantics of a sentence that is sensitive to sublexical components. Individual morphemes could be taken to contribute not only to the meaning of *Most1* vs. *Most2* but also to the interface with visual cognition.

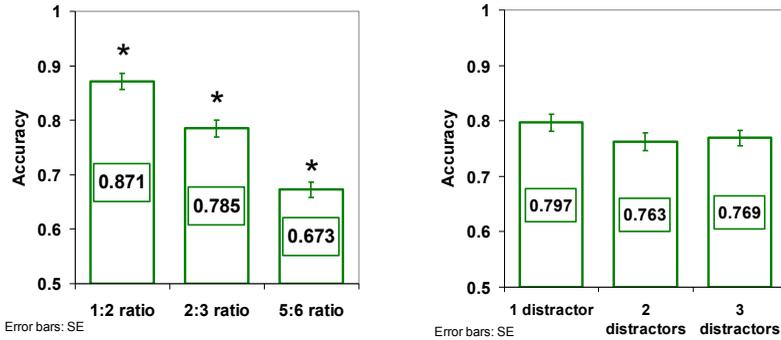
3.3 Results

3.3.1 *Most1* (Większość)

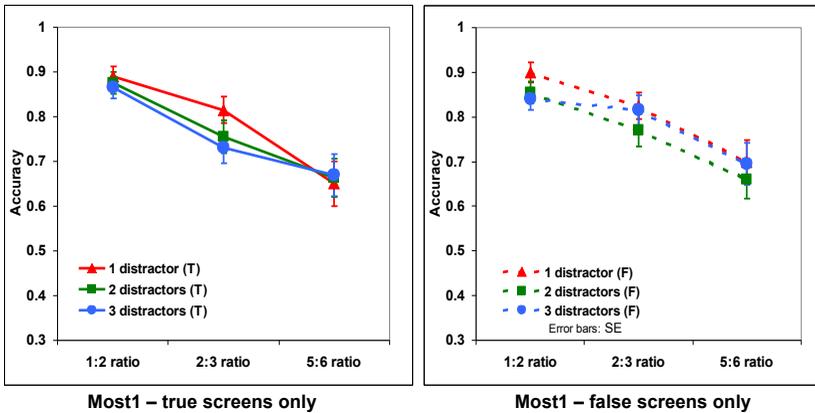
We conducted a 3x3x2 Repeated Measures ANOVA crossing the three levels of ratio and the three levels of number of distractor and truth/falsity of screens (i.e. whether ‘yes’ or ‘no’ is the correct answer). Our predictions were met – there was a significant effect of ratio ($F(2, 38) = 76.072, p < .001$), but no significant effect of distractor ($F(1.47, 27.98) = 1.637, p = .215$) (means can be seen in (10)). There were no significant interactions. The truth/falsity of screens with respect to the test sentence had no effect on the accuracy of participants’ judgments, which can be seen in (11).

The significant effect of ratio and no significant effect of distractor for *Most1* is the same as the findings for English *most* in Lidz et al. (2009). *Most1* is thus compatible with the Subtraction verification procedure in (2a). The selection of the target and the total is not affected by the number of distractor sets, but only by the ratio between the target set and the distractors.

(10) Accuracy of responses for *Most1*



(11) *Most1* on true and false screens



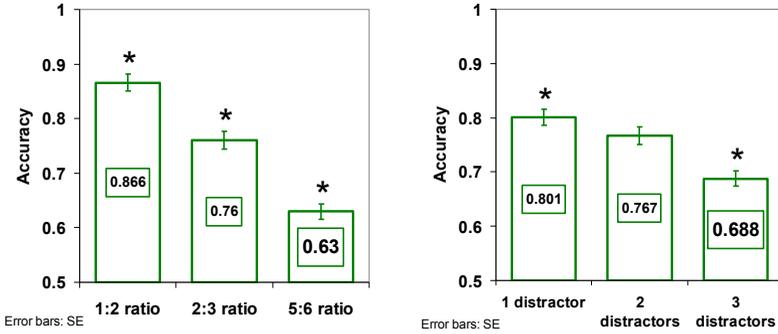
In the experiments of Pietroski et al. (2008) and Lidz et al. (2009), participants showed a bias towards a particular verification strategy for *most*, which resulted in a different pattern of accuracy than if a hypothesized alternative procedure had been used. We provided a control condition where an alternative verification procedure is necessary. We show that Subtraction continues to be used even on those conditions, where Selection can easily be performed and would in fact yield more accurate results.

3.3.2 *Most2 (Najwięcej)*

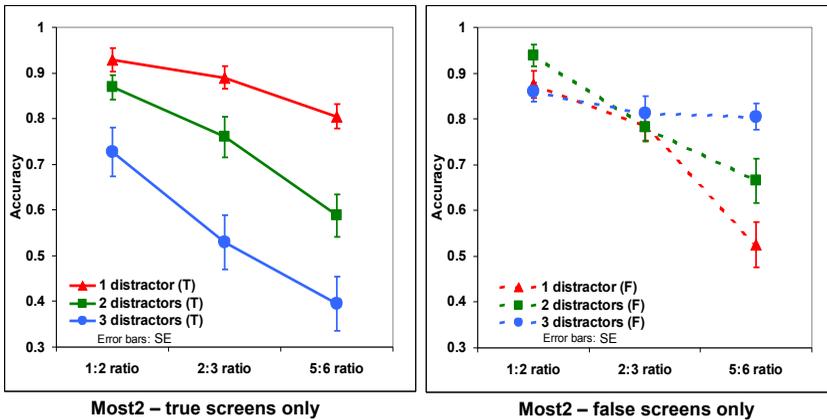
Our predictions were borne out: in addition to the effect of ratio $F(2, 38) = 124.77, p < .001$, there was a significant effect of distractor $F(2, 38) = 17.34,$

$p < .001$ (mean responses are in (12)). There was also a borderline significant interaction between ratio and distractor $F(4, 76) = 2.48, p = .051$.

(12) Accuracy of responses for *Most2*



(13) *Most2* on true and false screens



The graph in (13) shows a difference in accuracy patterns between true and false screens. There is no significant main effect of truth/falsity of screens (the overall mean for true screens .721, for false .783), so it is not the case that making a false judgment is easier. However, the significant interactions between distractor and truth/falsity, $p < .001$, and between ratio, distractor and truth/falsity, $p < .001$, indicate that participants made judgments

differently for true and false screens depending on the ratio and number of distractor colors.⁵

While accuracy rates with *Most1* were affected only by the ratio, accuracy rates with *Most2* were affected both by ratio and by the number of color sets. These results for *Most1* and *Most2* are consistent with the verification strategies in (2a) and (5), respectively. Since Subtraction (2a) does not depend on the number of distractor color sets, its computational cost remains the same as the number of distractors increases. Stepwise Selection (5), on the other hand, does become more computationally costly as the number of distractors increases.

Thus, *Most1* is not verified by Selection, as defined in (2b), although Selection is a psychologically plausible strategy given its use in (5). It is not psychophysics that forces the Subtraction strategy for *Most1* and English *most*. The instructions for the visual system are obtained directly from the relations expressed in the semantics. With *Most1* attention is never directed towards the individual distractor color sets, as predicted by the Interface Transparency Thesis. Further support for the thesis comes from the comparison of accuracy patterns on the condition when the screens for *Most1* and *Most2* were identical.

3.3.3 *Most2* vs. *Most1* on Two Color Screens

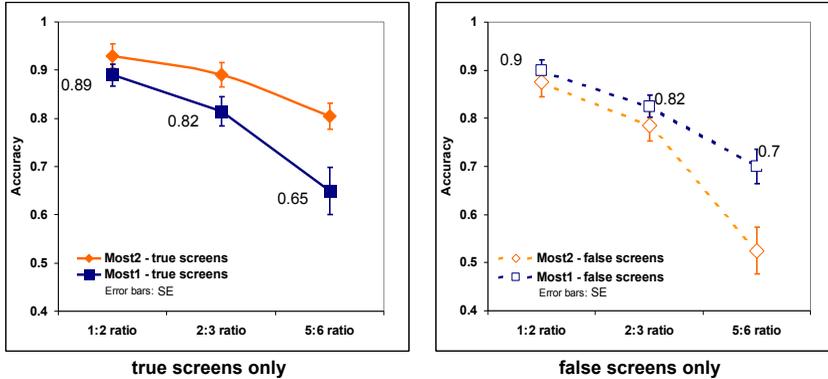
Lidz et al. (2009) argued for English *most* that the fact that accuracy was not greater on the two-color condition means that the information automatically computed by the visual system was not used. Therefore, the reason for the failure to directly select the comparison set must be the semantic representation of the sentence. Our results provide more direct evidence for this conclusion. Our participants behaved differently depending on which quantifier was used, even though the screens they judged were the same and either strategy would provide the correct judgment.

The patterns of accuracy for each quantifier were very different. In both cases the accuracy rates were a function of the ratio, but on true screens participants were significantly more accurate when selecting ‘yes’ with *Most2*. When selecting ‘no’ on false screens, they were more accurate with *Most1*.

⁵ This is not surprising since the Selection and comparison procedure is different at each step for the confirmation and disconfirmation of (8), i.e. *Yellow dots form the largest subset*. E.g., on a screen with 8 yellow, 10 blue, 7 green, 6 red dots, it is enough to notice that blues form the largest subset to disconfirm (8). With 8 yellow, 7 blue, 6 green, 5 red dots, comparison with each color set is needed to make sure that yellows are the most. What is surprising is the low accuracy with the disconfirmation of (8) on the two color screens. Apparently, judging that (8) is false with 8 yellow and 10 blue dots is harder than judging (8) true with 10 yellow and 8 blue dots.

Separate ANOVAs for true and false screens yield a significant main effect of quantifier type ($F(1, 19) = 10.49, p = .004$ for true screens and $F(1, 19) = 11.122, p = .003$ for false screens).

(14) *Most1* and *Most2* on the two-color condition



The different performance with each quantifier is fully predicted on the account that each involves a distinct verification strategy that is consistently used even when the screens are exactly the same. Different behavior (depending on the quantifier used) on the very same screens indicates not only that participants do not switch to a more efficient procedure, but that the way the procedures differ is specified by the semantics.

On screens with two color sets the computation for both *Most1* and *Most2* requires the comparison between the yellow and the non-yellow set. The instructions for how to perform this comparison are different for each quantifier, even though the components for the computation provided by the visual system are the same: the yellow set, the non-yellow set, the superset.

The accuracy with *Most1* was no different on true and false screens; the means for each ratio were almost identical. This result is fully predicted by Subtraction – the same computation is performed for both positive and negative judgments: e.g., with 8 yellow and 7 blue dots (true) and 8 yellow and 10 blue dots (false) (i) estimate the target yellow set, 8, (ii) estimate the total, 15 (true) or 18 (false), (iii) subtract the target from the total 15-8 or 18-8, (iv) compare the cardinalities from (i) and (iii) $8 > 7$ (true) or $8 > 10$ (false).

With *Most2*, in order to confirm that yellow dots form the larger of the two sets, the non-yellow set is selected directly. This results in higher accuracy than confirmation with *Most1* where the non-yellow set is computed. A ‘no’ judgment with *Most2*, however, results in significantly more

errors than with *Most1*, e.g., on a screen with 8 yellow and 10 blue dots.⁶ Despite this puzzling effect of accuracy drop with *Most2* on false screens, it is clear that each quantifier relies on a dedicated strategy for verification. Participants do not switch to the more advantageous strategy (e.g., they do not use Selection to more accurately confirm the truth of sentences with *Most1*, or Subtraction to more accurately disconfirm the truth of sentences with *Most2*). The two distinct accuracy patterns provide strong evidence that the lexical meaning of the functional morphemes that build up *Most1* and *Most2*, and their logical syntax, are interfacing with the visual cognition during the verification process.

4 Conclusions

Our experiments indicate that semantics provides a direct set of instructions to visual cognition processes. Each of the two Polish quantificational expressions *większość* (*Most1*) and *najwięcej* (*Most2*) biases a particular verification strategy that is used as a default under time pressure. Each canonical verification strategy overrides other cognitively available strategies for truth verification as proposed in Lidz et al. (2009). The following predictions were met:

(i) Polish *Most1*, like English *most*, is verified using the Subtraction strategy. The accuracy in the verification of a sentence containing *Most1* is sensitive to (i) the ratio between the cardinality of the target color set and (ii) the cardinality of the set of dots in other colors. Response accuracy was unaffected by the number of distractor color sets. The significant effect of ratio and no effect of the number of distractors with Polish *Most1* directly replicate the findings of Lidz et al. (2009) for English *most*.

(ii) A closely related quantifier *Most2* requires the Stepwise Selection strategy. The response accuracy with *Most2* depends on both the ratio and the number of distractors. The availability of Stepwise Selection with *Most2* indicates that it is not psychophysics that precludes the use of Selection with *Most1* and English *most*. This result provides direct evidence for the Interface Transparency Thesis put forth by Lidz et al. (2009), according to which verification procedures bias those algorithms that directly compute the semantic representation.

⁶ This result could be related to the so-called “confirmation bias” observed in psychology (Nickerson 1998), so that participants were more likely to overestimate the yellow set and underestimate the non-yellow set. On the 5:6 ratio condition the difference between the yellow and the non-yellow set was only 1-2 dots.

(iii) Our results also suggest that each verification strategy is canonical in that it is followed consistently for each lexical item. Specifically, this is indicated by the finding that the same group of participants behaved differently depending on the quantifier. On the two color condition where sentences with *Most1* and *Most2* were either both true or both false, participants did not switch to the more effective strategy; rather, the properties of the linguistic input directly influenced the unconscious decision making system associated with visual cognition.

Importantly, the results confirm and extend the proposals and findings of Pietroski et al. (2008), Hackl (2009), Lidz et al. (2009) that the compositional semantics of quantifiers interacts in predictable ways with the visual system during verification.

References

- Dehaene, Stanislas. 1997. *The number sense: How the mind creates mathematics*. New York: Oxford University Press.
- Feigenson, Lisa, Stanislas Dehaene & Elizabeth Spelke. 2004. Core systems of number. *Trends in Cognitive Sciences* 8(7). 307–314.
- Hackl, Martin. 2009. On the grammar and processing of proportional quantifiers: *most* versus *more than half*. *Natural Language Semantics* 17(1). 63–98.
- Halberda, Justin, Len Taing & Jeffrey Lidz. 2008. The development of “most” comprehension and its potential dependence on counting-ability in preschoolers. *Language Learning and Development* 4(2). 99–121.
- Halberda, Justin, Sean F. Sires & Lisa Feigenson. 2006. Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science* 17. 572–576.
- Lidz, Jeffrey, Paul Pietroski, Tim Hunter & Justin Halberda. 2009. Interface Transparency and the Psychosemantics of *most*. Article in press. *Natural Language Semantics*.
- Nickerson, Raymond S. 1998. Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology* 2. 175–220.
- Pietroski, Paul, Jeffrey Lidz, Tim Hunter & Justin Halberda. 2008. The meaning of *most*: Semantics, numerosity and psychology. *Mind and Language*, 24(5). 554–585.
- Trick, Lana M. 2008. More than superstition: Differential effects of featural heterogeneity and change on subitizing and counting. *Perception and Psychophysics* 70(5). 743–760.

- Whalen, John, C.R. Gallistel & Rochel Gelman. 1999. Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science* 10(2). 130–137.
- Wiese, Heike. 2003. *Numbers, Language, and the Human Mind*. Cambridge: Cambridge University Press