Chapter 17

Some, *most*, *all* in a visual world study

Barbara Tomaszewicz-Özakın
University of Cologne

In a visual world eye-tracking study I find that Polish quantifiers *niektóre* 'some', *większość* 'most of', *najwięcej* 'the most' and *wszystkie* 'all' elicit distinctive patterns of looks, consistent with their semantics. *Niektóre* 'some' has a strong scalar implicature: the meaning 'some-not-all' is processed immediately as the quantifier is heard. The superlative *najwięcej* 'the most' quickly triggers comparisons between the target and the other sets. The proportional *większość* 'most of' elicits a pattern suggesting that its verification involves the estimation of the total set. With *wszystkie* 'all' the identification of the target set is the fastest.

Keywords: quantifiers, semantics–cognition interface, eye-tracking, Polish

1 Introduction: Interpreting quantifiers

We talk about quantities all the time while describing the world. Quantifiers are natural language expressions used to describe quantities with or without using number terms. Semantically they express relations between sets (Barwise & Cooper 1988), e.g., *most of* in the sentence *Most of the balls are blue* tells us that the set of balls that are blue is larger than the set of non-blue balls. We can easily assess the conditions that make this sentence true/false, but how do we verify such sentences in real-life situations? The generalized quantifier theory (Mostowski 1957, Lindström 1966, Montague 1973) is silent about this issue, mainly because for philosophers (Montague 1973, a.o.) semantics was a branch of mathematics and not of psychology (Partee 2011). Psychologists, however, have long been studying the number sense in humans, a dedicated brain system for abstract representation of number and the source of our mathematical intuitions, which is employed in the judgments involving quantifiers (Feigenson et al. 2004,

Barbara Tomaszewicz-Özakın. 2021. *Some*, *most*, *all* in a visual world study. In Mojmír Dočekal & Marcin Wągiel (eds.), *Formal approaches to number in Slavic and beyond*, 399–426. Berlin: Language Science Press. DOI: 10.5281/zenodo.5082482

McMillan et al. 2005, 2006, Clark & Grossman 2007, Dehaene 2009, 2011, Troiani et al. 2009). Speakers have different ways of referring to quantities: number terms when the specific size of the set is at issue, and vague quantifiers like *some* and *most* or context-dependent quantifiers like *few* and *many*, when they refer to approximate quantities. The present study addresses two critical questions about quantifiers: (i) what is included in the lexical representation of quantifier meanings and (ii) what the psychological mechanisms involved in the interpretation of those meanings are.

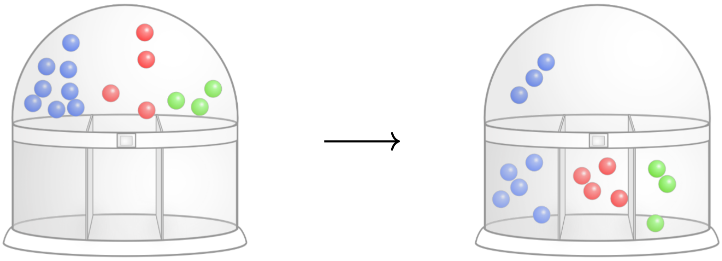
Investigating those two questions, Pietroski et al. (2009) and Lidz et al. (2011) put forth a novel hypothesis that what participants are doing to verify sentences containing quantifiers, i.e., (ii), can be directly determined by quantifier semantics, (i). To illustrate, in order to verify that most balls are blue we need to compare the numbers of blue and and non-blue balls, but how do we obtain the number of non-blue balls? If there are balls in one other color, we can simply count them. If there are more colors, we can count the numbers of balls in each other color and add them up; or we can obtain the number of all balls and subtract the number of the blue balls from it; or we can instead verify if blue balls are more than the half of all balls. But do we even need to count? Children who are not yet able to count are perfectly able to understand sentences containing *most* (Halberda et al. 2008, Odic et al. 2018), and in real-life situations we do not need to know precise quantities to use *most*.

Pietroski et al. (2009), Bates, Kliegl, et al. (2015), Tomaszewicz (2011, 2012, 2013, 2018), Hunter et al. (2017), Knowlton et al. (2021) obtained experimental evidence that *most of* induces a subconscious choice of a procedure based on subtraction for verification against visual displays, even in situations where comparing the numerosity of the target set and one other set directly would be more efficient. Why would the mind not subconsciously choose the most efficient procedure in a given situation? According to the hypothesis it is because the mind follows the "instructions" encoded in the logical function representing the meaning of a quantifier. Tomaszewicz (2011, 2012, 2013, 2018) showed that, in contrast to the proportional quantifier *most of* (Polish *większość*), the superlative *najwięcej* 'the most', as in *Najwięcej kulek jest niebieskich* 'Blue balls are more numerous than balls in any other color', directs participants' subconscious attention to obtaining the numerosities of each other color set. The participants were prompted to switch between verification procedures by a change in the linguistic input, but not by a change in the visual input. Thus, the motivation for the subconscious switch in procedures is not to maximize efficiency. Participants used the procedure associated with each quantifier, and in effect, the same display was verified differently depending on which information the visual system was instructed to use by the lexical representation of quantifier meanings.

The present experiment was designed to uncover the details of the lexical semantic specification of the Polish quantifiers *niektóre* 'some' and *większość* 'most of' in comparison to *wszystkie* 'all' and *najwięcej* 'the most'. It utilizes eye-movement as a representational measure in the visual world paradigm. Visual world eye-tracking has been used to demonstrate how comprehenders rapidly integrate different sources of information in order to identify the referents in the visual display as the sentence unfolds over time (Tanenhaus et al. 1995, Allopenna et al. 1998). The visual world paradigm allows for the closest approximation of real-life visual contexts in an experimental setting. Our conscious experience is that our eyes glide from one thing to another thing, but, in fact, unless we are tracking a moving object, our eyes perceive images in a series of rapid jerky movements (saccades). We can track the series of fixations at a particular point and the saccades away from that point in order to analyze which parts of the image attract attention and how. In a visual world task, participants hear the sentence as they inspect the visual scene and their eye movements are recorded; in particular, the proportion of looks to the target in the picture is measured. This makes it an excellent tool for the investigation of incremental processing. At the point in the sentence when the quantifier is heard, participants' subconscious attention should be directed to different aspects of the scene, towards or away from the target, depending on the semantics of the quantifier.

In the current experiment, the four Polish quantifiers, *wszystkie* 'all', *niektóre* 'some', *większość* 'most of' (proportional *most*), and *najwięcej* 'the most' (superlative *most*, henceforth *most-SUP*), appeared in the same carrier sentence, describing the same identical display for *some*, *most of* and *most-SUP* (screens for *all* needed to differ as will be explained shortly). I employed the gumball paradigm of Degen & Tanenhaus (2011, 2016). Participants evaluated sentences of the form 'You got all/some/most of/most-SUP blue balls' against displays of a ball machine dispensing balls of three colors from upper to lower chambers; see Figure 1. The correlate of the processing of the information about the quantifier semantics was the proportion of looks to the target (the set of blue balls) vs. the so-called distractors (the two other color sets).

I thus build on the results of Huang & Snedeker (2009, 2011), Grodner et al. (2010), Degen & Tanenhaus (2011, 2016), who showed that contextual effects on the interpretation of quantifiers are reflected in eye-movement patterns. Those studies investigated the time course of scalar implicatures, i.e., the pragmatic aspects of the meaning of the quantifier *some*, while I concentrate on the precise semantic distinctions between the four Polish quantifiers to test both the scalar implicature of Polish *niektóre* 'some' (Spychalska 2009) and the verification procedure associated with Polish *większość* 'most of' (Szymanik & Zajenkowski



The ball machine at the beginning of the trial

After a button press, the machine dispenses balls to the lower chambers and after 500 ms the stimulus sentence is played

Figure 1: The blue set is the target.

2010). The results of the visual verification experiments (visual search paradigm) in Pietroski et al. (2009), Lidz et al. (2011), Tomaszewicz (2011, 2012, 2013), Hunter et al. (2017), and Knowlton et al. (2021) indicate that quantifier semantics guides the subconscious adoption of a verification strategy. The visual search paradigm involves comparisons of the accuracy of judgments with reference to rapidly presented displays (200–300 ms), i.e., accuracy is taken as a proxy for the processing cost. In these prior experiments the number of different color sets affected accuracy in different ways depending on the quantifier in the stimulus sentence. So while in the visual search paradigm it is assumed that the sentence stimulus somehow provides an instruction for verification, the visual world paradigm in the current study enables us to tap into the real-time construction of this instruction as the auditory stimulus unfolds.[1]

2 The current study

2.1 Methods

In each trial (72 trials in 3 blocks of 24), participants ($n = 35$), saw a fixation cross and the display of a ball dispensing machine with its upper chambers filled with 3 colors of balls (bottom chambers empty), as in the left panel of Figure 1.

[1]See Huettig et al. (2011) for an argument how the two paradigms, visual world and visual search, provide converging evidence for the role of working memory in the interactions between linguistic input and visual attention.

After 2 seconds, the button in the center of the machine turned yellow, and the participants clicked on the button. Upon clicking, a grey mask was displayed for 200 ms. (Clicking the central button ensured that the participants were looking at the central fixation point at the time of the auditory stimulus onset.) Now the second display was shown: the ball machine was redisplayed with a certain number of balls of each color having dropped to the lower chamber, e.g., right panel in Figure 1.

After 500 ms, the participants heard one of the stimulus sentences in (1). Their task was to click on that lower chamber which contained the balls mentioned in the statement if they thought the statement was true, and click on the central button otherwise.

(1) Dostałeś…
got.PAST.2SG
'You got…'

a. większość {niebieskich / czerwonych / zielonych} kulek. MOST-OF
most.of blue red green balls
'most of the blue/red/green balls.'

b. najwięcej {niebieskich / czerwonych / zielonych} kulek. MOST_{SUP}
most-SUP blue red green balls
'the most blue/red/green balls.'

c. niektóre {niebieskie czerwone zielone} kulki. SOME
some blue red green balls
'some blue/red/green balls.'

d. wszystkie {niebieskie / czerwone / zielone} kulki. ALL
some blue red green balls
'all of the blue/red/green balls.'

All the sound files were cross-spliced and normalized using Praat Vocal Toolkit (Corretge 2020) so that all the quantifiers and color expressions had the same duration. Once the participants clicked indicating their response, a grey screen was displayed for 1s and the experiment advanced to the next trial. Participants' eye movements were recorded with an Eyelink 1000 eye-tracker at a sampling rate of 1000 Hz.

There were 8 conditions: 4 quantifiers (SOME/MOST-OF/MOST_{SUP}/ALL) * 2 display types (EARLY/LATE). In EARLY trials there was only one partitioned set, e.g. the blue set in Figure 1. In LATE trials all sets were partitioned. The difference between the EARLY and LATE displays is discussed and illustrated with pictures in the next section (Figure 4). The displays for the test sentences for the analysis of eye-movements required a Yes response. Filler trials, half of all trials, required a No response. For Yes responses participants clicked the chamber that matched the sentence, for No responses they clicked the button. Following Degen & Tanenhaus (2011, 2016) I included what they call GARDEN-PATH trials among the fillers in order to force the participants to pay attention and notice that the sentences throughout the experiment might not always be true. On these trials, one set was partitioned allowing an anticipation of a quantifier, but as the sentence unfolded it turned out this set did not match the sentence (leading to a garden-path-like effect where you had to revise your search for the target); see Figure 2.

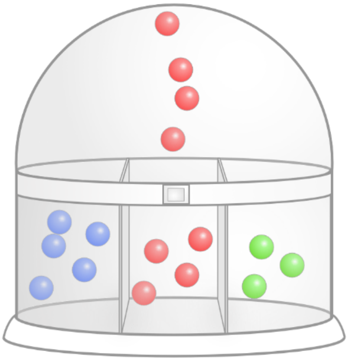
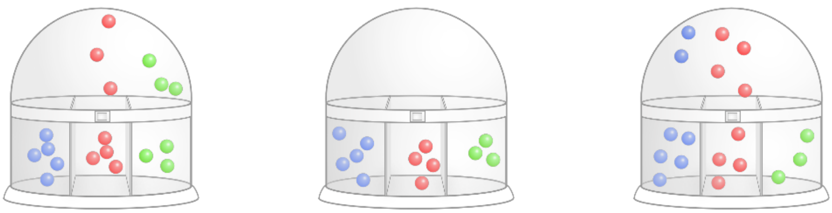


Figure 2: GARDEN-PATH condition

The displays were identical for 3 quantifiers: *some*, *most of*, *most*-SUP. The quantifier *all* had different displays, because the top chamber for that color had to be empty, Figure 3.



all-EARLY *all*-LATE *all*-GARDEN-PATH

Figure 3: The displays for the ALL condition

2.2 Predictions

The visual world studies in Huang & Snedeker (2009, 2011), Grodner et al. (2010), and Degen & Tanenhaus (2011, 2016) compared the quantifiers *some* and *all* in order to establish whether the processing of the scalar implicature of *some* is delayed. The literal interpretation of *some*, as in *You got some of the blue balls*, is that you got at least one blue ball, so if you got all of them, the sentence is true.[2] But in most contexts, we understand this sentence as saying that we got some blue balls but not all of them. This interpretation is a pragmatic inference: the speaker would have said 'You got all of the blue balls' if we got all of them because that would be a more informative statement. Quantifiers *all* and *some* form a scale, so when *some* is used instead of the stronger *all*, the meaning 'some-not-all' is inferred – this inference is called a scalar implicature (Horn 1972, Levinson 1983, a.o.). Huang & Snedeker (2009) found that the 'some-not-all' reading was delayed in comparison with *all*, but Grodner et al. (2010) found no delay and Degen & Tanenhaus (2011, 2016) hypothesized that the reason for the delay in Huang & Snedeker (2009) was the availability of other descriptions of the scene: sentences with number terms in addition to *some* and *all*. Degen & Tanenhaus (2016) indeed found that when no such alternatives were present, the processing of the *some-not-all*-implicature was not delayed relative to the processing of the meaning of *all*, but it was somewhat delayed when those alternatives were available. Thus, the processing of pragmatic meaning may be no more costly than the processing of the literal meaning of a quantifier, depending on the context. In the present experiment, I used the gumball paradigm of Degen & Tanenhaus (2011, 2016) to investigate the time course of processing of both pragmatic and semantic information.

Given the findings of Degen & Tanenhaus (2011, 2016), the implicature of the Polish 'some', *niektóre*, could be delayed due to the presence of alternative utterances that could describe the same situation. However, Spychalska (2009) argues that the implicature of *niektóre* is stronger than that of English *some*, so if we find no delay, the current methodology is a useful tool for the investigation of cross-linguistic semantic differences. To get more specific information about the time course of the processing of *niektóre*, it is compared with the two majority quantifiers whose literal meaning allows us to make precise predictions for processing, given the findings in Tomaszewicz (2011, 2012, 2013, 2018) that each of them drives a distinct verification procedure consistent with its semantics.

[2] As observed in Spychalska (2009) in Polish *niektóre* must mean 'at least two', because the quantifier occurs only in the plural form.

The superlative *most* (*most*-sup) in the sentence *You got the most blue balls* (true in the right top panel of Figure 4) requires a comparison between the balls in the lower chambers of the machine. In Figure 4, the machine has dropped down 4 blue, 3 red, 2 green balls. Figure 4 illustrates how the stimulus sentence unfolds: when hearing the quantifier *najwięcej* (*most*-sup, i.e., 'the most'), you already know you need to compare the numbers of the balls and if you had already determined that the blue set is the biggest, you can anticipate the adjective 'blue'. The proportional quantifier, *most of*, on the other hand, requires you to compare the numbers of the blue balls in the lower and the upper chambers. So these two majority quantifiers require very distinct patterns of eye-movements. The proportion of the looks to the target blue set at the moment of hearing the quantifier should be lower with *most*-sup because two comparisons are needed (with the red and the green set) whereas with *most of* just one comparison is necessary (between the lower and upper blue sets). This predicted contrast between *most of* and *most*-sup allows us to test whether the looks to the target with *some* will be delayed like with the *most*-sup. In addition, obtaining this contrast would set up a baseline for further visual world studies on majority quantifiers cross-linguistically.

The experiment also includes a comparison with the universal quantifier *all*, but it was not possible to present it together with the same displays as for the three other quantifiers; see the bottom panels of Figure 3. The displays for *all* contained the same numbers of balls in the lower chambers and the colors were in the same order as in the corresponding displays for the other quantifiers (the exact location of the balls within a chamber was a little different because the displays were generated with a random scatter). The displays for *all* make it very easy to anticipate the color adjective at the point of hearing the quantifier so this condition provides us with a time course for the highest proportion of looks to the target (I already note here that this is not the baseline for statistical comparisons because I want to compare the quantifiers with the same identical displays).

The contrasts described above are predicted for the displays where the quantifier provides a point of disambiguation as to which color is the target set. This is the early condition, i.e., in these displays target identification can happen earlier than the color adjective is heard. The looks to the target set in the early condition should begin to increase in the quantifier window (as in Degen & Tanenhaus 2016). In contrast, in the LATE condition, see Figure 5 below, the point of disambiguation is the color adjective.

The theoretical predictions outlined above may be affected by possible confounds, because in experiments on visual identification participants may exhibit

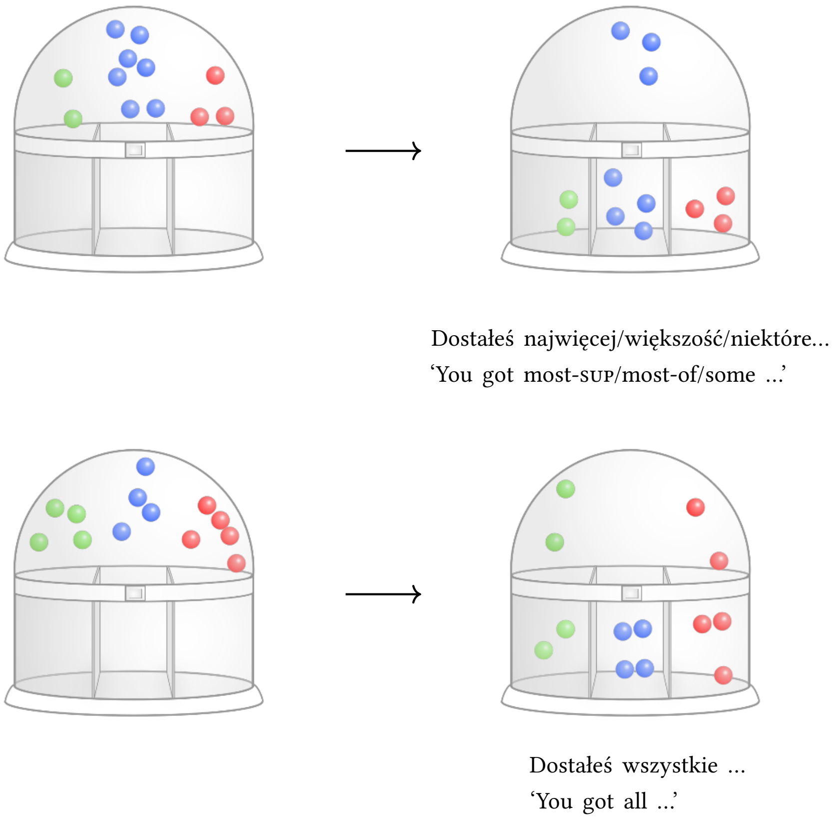


Figure 4: Sample displays in EARLY condition – target: blue.

different kinds of biases that come, for instance, from the way the visual system works. One of them is the bias to look more at larger set sizes (also found in the study of Degen & Tanenhaus (2016). This means that already during the preview of the picture, before the sentence is heard, participants will tend to look at the blue set in the top right panel of Figure 4. We also know that when precise counting is impossible or simply not needed as in the current experiment, people use the Approximate Number System (ANS) that generates a representation of magnitude rather than an exact cardinality, (Feigenson et al. 2004, Dehaene 2009, 2011). It is also known that with a 500 ms display ANS automatically enumerates the total set (the superset) and up to two color subsets in parallel, (Halberda et al. 2006). Thus, the time course of eye-movements over the three regions of interest is expected to reflect the following effects in the EARLY condition (the summary is in Table 1).

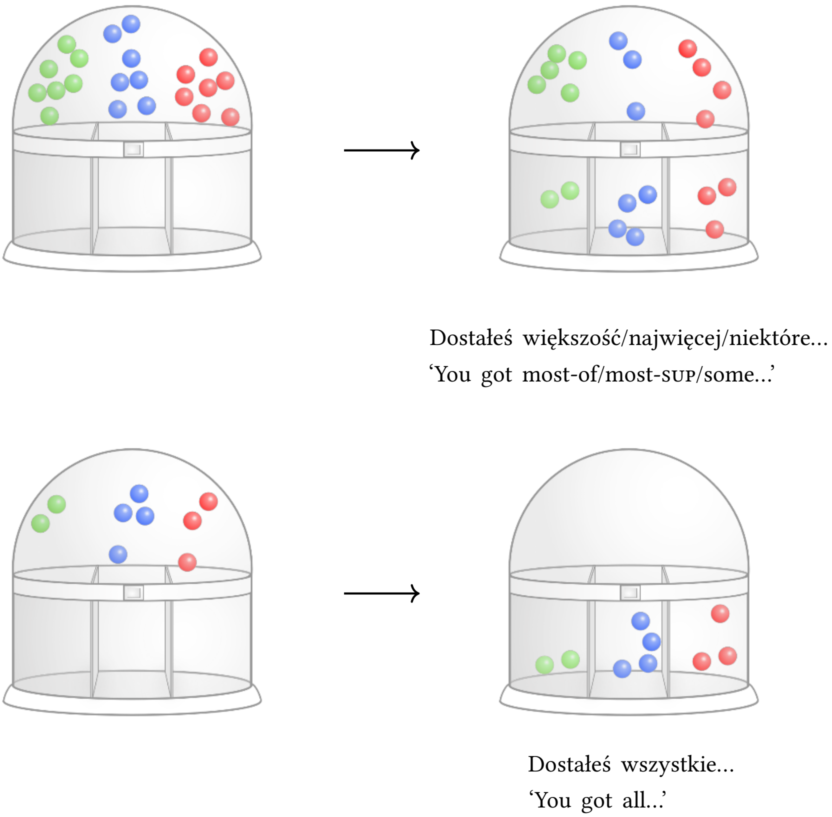


Figure 5: Sample displays in LATE condition – target: blue.

2.2.1 Preview and 'You got' (EARLY)

During the 500 ms preview of the display (the right panel of Figures 1–3) and during the beginning of the sentence (*Dostałeś...* 'You got...') I expect no differences in the looks to the target for the four quantifiers, except for the bias to look at the biggest set (the blue target pops out as different than the other two sets, hence it may attract looks early on).

2.2.2 Quantifier (EARLY)

In the EARLY condition, where the quantifier disambiguates which set is the target, in the quantifier window, I expect fewer looks to the target with *most*-SUP than with *most of* and with *all*. The theoretical prediction explained above is that *most*-sup requires two comparisons (between the blue and the red set, and between the blue and the green set in the lower chambers), while *most-of* requires one (between the blue set in the lower chamber and that in the upper chamber). During the preview and the ‘You got’ window the looks may be attracted to the biggest and partitioned set which is the blue target, therefore in the quantifier window the looks may already move to the distractors. The quantifier *all* requires no comparisons. This prediction is summarized as ‘most_{sup} < most-of, all’ in Table 1 (‘fewer looks to the target set in the most_{sup} condition than in the most-of and all conditions’).

The predictions for *some* in the early condition in the quantifier window depend on which interpretation could be in the minds of the participants at this point. If the scalar implicature, ‘some-but-not-all’, has already been processed, the identification of the target should be (almost) just as easy as with *all*: it cannot be the red nor the green set, and the blue partitioned set has already stood out during preview. Hence, ‘some-not-all = all’ in Table 1. This interpretation would also attract more looks to the blue target than with *most*-sup, ‘some-not-all > most_{sup}’. If, instead, participants are first considering the literal meaning of *some*, ‘some-and-possibly-all’, then their looks will be directed to the green and red color sets as with *most*-sup, ‘some-possibly-all = most_{sup}’.

Finally, there should be more looks to the target in the early condition in the quantifier window with *some* on the ‘some-but-not-all’ interpretation than with *most of*, ‘some-not-all > most-of’. With both quantifiers the looks will be attracted to the partitioned set, the blue target, but with *most of* you need to estimate the numerosities of the two blue subsets and compare them to verify that the sentence is true.

2.2.3 Color + ‘balls’ (early)

In the early condition, all but one of the effects observed in the quantifier window are predicted to be carried over to the color window. The exception is the quantifier *most of*, which now may attract a similar proportion of looks to the target as *some-not-all*, ‘some-not-all = most-of’. The alternative is that with *most of* there will still be more looks away from the target, ‘some-not-all ‘ most-of’, because ‘*most of the blue balls’* requires more operations for visual verification than the comparison of the top and bottom blue set as I stated above. Pietroski et al. (2009) and Lidz et al. (2011) propose that sentences with *most of* are verified against visual displays of multicolored dots not by directly comparing two sets but by a subtraction procedure: you estimate the superset, you estimate the target set, subtract and compare the result with the target. This procedure involves more steps than direct comparison of two sets but the reason it is followed is because it is directly specified in the lexical semantics of the proportional quantifier *most*. Lidz et al. (2011) argue that sentential meanings are "individuated more finely than truth conditions" (p. 2) precisely because they interface with perception systems such as visual cognition. It has been established that numbers can be represented as "noisy magnitudes" even for the purposes of basic arithmetic operations like addition and subtraction (Feigenson et al. 2004, Degen & Tanenhaus 2011), so the subtraction procedure is possible even with a 200 ms display, but crucially it is less efficient than direct comparison. This effect was shown in the visual search studies of Pietroski et al. (2009), Lidz et al. (2011), Tomaszewicz (2011, 2012, 2013), Hunter et al. (2017), Knowlton et al. (2021), which measured accuracy of Yes-No responses. In the current study I should find evidence that participants follow the subtraction procedure, as specified in (2), in contrast to direct selection of the two sets as in (3), if we find fewer looks to the target in the color window than with SOME-NOT-ALL because of continuing looks to the top blue set in order to establish the total set. Perhaps, the proportion of looks to the target will even be as low as with *most*-SUP ('SOME-NOT-ALL = MOST-OF?/SOME-NOT-ALL > MOST-OF/MOST-OF = MOST_{SUP}?' in Table 1). Such a result in the color window in the EARLY condition would provide support for a higher number of processing steps involved in (2) as opposed to (3).

(2) SUBTRACTION procedure for the verification of the sentence 'You got *most of* the blue balls':
#[BLUE(x) & BELOW(x)] > #[BLUE(x) & ABOVE(x) & BELOW(x)] − [BLUE(x) & ABOVE(x)]

(3) SELECTION procedure for the verification of the sentence 'You got *most of* the blue balls':
#[BLUE(x) & BELOW(x)] > #[BLUE(x) & ABOVE(x)]

The differences expected to occur in the LATE condition are presented in the following subsections.

2.2.4 Preview and 'You got' (LATE)

I expect no differences. As can be seen in Figure 5, bottom-right panel, the target set cannot be identified during the preview by the big set bias (the blue bottom set is not the only large set).

Table 1: Predictions. (†) marks the *point of disambiguation*.

(a) EARLY

Preview	'You got'	Quantifier (†)	Color + 'balls'
No differences/ Big set bias?		MOST$_{\text{SUP}}$ < MOST-OF & ALL SOME-NOT-ALL = ALL SOME-NOT-ALL > MOST$_{\text{SUP}}$ SOME-POSSIBLY-ALL = MOST$_{\text{SUP}}$ SOME-NOT-ALL > MOST-OF	MOST$_{\text{SUP}}$ < MOST-OF & ALL SOME-NOT-ALL = ALL SOME-NOT-ALL > MOST$_{\text{SUP}}$ SOME-POSSIBLY-ALL = MOST$_{\text{SUP}}$ SOME-NOT-ALL = MOST-OF? / SOME-NOT-ALL > MOST-OF / MOST-OF = MOST$_{\text{SUP}}$?

(b) LATE

Preview	'You got'	Quantifier	Color + 'balls' (†)
No differences		MOST$_{\text{SUP}}$ > ALL/MOST-OF/SOME SOME-NOT-ALL > MOST-OF SOME-NOT-ALL < ALL	MOST$_{\text{SUP}}$ < ALL/MOST-OF/SOME SOME-NOT-ALL > MOST-OF SOME-NOT-ALL < ALL

2.2.5 Quantifier (LATE)

In the LATE condition, the target set can only be reliably disambiguated upon hearing the color adjective, that is, in the last time window of interest. However, I do expect differences in the quantifier window already.

Because the target set cannot be biased during the preview, upon hearing the quantifier *most*-SUP, the looks could be immediately directed to the largest of the bottom sets, the blue target, while with *most of* and *some* the looks will also be directed to the upper sets and with *all* to the other bottom sets. Thus, the prediction is 'MOST$_{\text{SUP}}$ > ALL, MOST-OF, SOME' in Table 1. Alternatively, the identification of the largest set with *most*-SUP is delayed until the color window, but given the big set bias, I find this option unlikely.

I also expect more looks to the target with *some-not-all* than *most of* because *most of* requires the estimation of the numerosity of the bottom blue set relative to the top set (in one of the two ways in (2–3) discussed above). Additionally, there should be fewer looks to the target with *some-not-all* than with *all* because the set for the latter is unpartitioned. These two effects should persist in the color window.

In the LATE condition, the *some-possibly-all* interpretation is not tested because all sets are partitioned.

2.2.6 Color + ‘balls’ (LATE)

At the point of hearing the color adjective in the LATE condition, the proportion of looks to the target with *most*-SUP should be lower than with other quantifiers because now the looks are attracted to the other two color sets in order to make the comparisons to confirm that indeed the blue set is the largest, ‘MOST_{SUP} < ALL, MOST-OF, SOME’. Could it be that once the largest set is identified already in the quantifier window, participants stop making the comparisons upon hearing ‘blue’ because it matches the already identified target? I do not think so, simply because the ‘identification of the target’ as early as the quantifier happens unconsciously, and only when the color is heard are the participants aware of the semantics of the full sentence, thus I expect the processing to keep going and to follow the semantics of the superlative sentence: ‘There are more blue balls than the balls in any other color’. Accordingly, I expect that in the color window, comparisons with other colors will take place.

Of all of the above, the predictions of main theoretical interest are the following:

(i) In the EARLY condition, at the quantifier (which disambiguates the target set) there will be fewer looks to the target with MOST_{SUP} than MOST-OF and ALL because the superlative semantics requires comparisons with other color sets, MOST_{SUP} < MOST-OF & ALL. The looks to the target with MOST_{SUP} can serve as the baseline for establishing if the implicature of *niektóre* ‘some’ is processed early: SOME-NOT-ALL > MOST_{SUP} vs. SOME-POSSIBLY-ALL = MOST_{SUP}.

(ii) In the EARLY condition, in the last region (Color + ‘balls’), with MOST-OF the looks will either stay on the target as with SOME (if participants follow the direct Selection procedure in (3)), SOME-NOT-ALL = MOST-OF, or there will be fewer looks to the target (if participants need to establish the total set of blue balls for the Subtraction procedure in (2)), SOME-NOT-ALL > MOST-OF, MOST-OF = MOST_{SUP}.

(iii) In the LATE condition, at the quantifier, there should be more looks to the target with MOST_{SUP} than with the other quantifiers, reflecting the immediate processing of the superlative semantics, MOST_{SUP} > ALL, MOST-OF, SOME.

(iv) In the LATE condition, at the disambiguation point (Color + ‘balls’), the looks to the target with MOST_{SUP} should decline, MOST_{SUP} < all, MOST-OF, SOME, because the semantics requires comparisons with other color sets.

2.3 Results: Behavioral

The mean accuracy on the test conditions (i.e, EARLY and LATE that required a Yes response) was 95%. Of the 35 participants, 30 got 97–100% correct and 3 got less than 70% correct (54%, 58%, 65%). I kept all of the responses because I did not aggregate the data for statistical analyses and I used the eye-movement data only from the correct trials. I removed the extremely long outlier reaction times (three standard deviations above the mean); those constituted 1.3% of the Yes and No data and were equally found in all conditions and regions of interest. The accuracy of the responses and reaction times (RTs) are plotted in Figure 6.

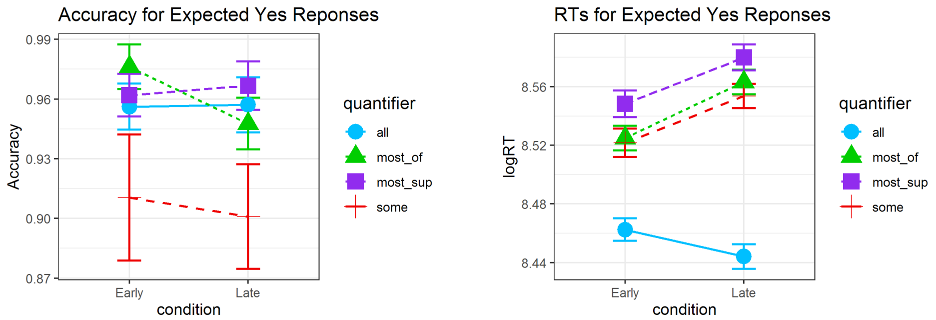


Figure 6: Accuracy and log-transformed reaction times (error bars represent standard errors)

I fitted a mixed-effects regression model of the log-transformed RTs and a mixed effects logistic regression model of the (binary) yes/no variable in R version 3.6.2 (R Core Team 2017) using the lme4 package version 1.1-21 Bates, Mächler, et al. (2015). The p-values were obtained using model comparison and the Satterthwaite approximation implemented in the lmerTest package (Kuznetsova et al. 2017).

For the accuracy data, there is a significant main effect of quantifier ($\chi^2 = 13.72$, df = 3, $p = 0.003$). With ANOVA-style contrast coding, there are no differences in pairwise comparisons between the conditions. There are also no differences in pairwise comparisons when ALL is the baseline; with SOME as the baseline only MOST-OF is significantly different, $\beta = 2.066$, SE $= 0.758$, $t = 2.742$, $p = 0.026$ (including the Bonferroni correction for multiple comparisons). Summing up, the accuracy across the conditions ranged from 87% to 99%, and participants were significantly less accurate in the SOME condition in comparison to the MOST-OF condition. We do not see such a difference in reaction times (right panel in Figure 6): SOME is not slower than the other conditions, which means that this condition was not harder, but either that people were fast and made mistakes (which is unlikely given that MOST-OF and MOST_{SUP} had similar RTs) or rather that they believed 'No, I did not get *some* of the balls in color X, I got *most* of them.'

The plot of the RTs in Figure 6 shows significant effects of the EARLY/LATE condition ($\chi^2 = 9.39, \text{df} = 1, p = 0.002$) and Quantifier ($\chi^2 = 72.91, \text{df} = 3, p < 0.0001$) and their interaction ($\chi^2 = 11.62, \text{df} = 3, p = 0.009$). Pairwise-comparisons with MOST_{SUP} as the baseline confirm what we see in the plot: that only the ALL condition is significantly faster ($\beta = -0.085, \text{SE} = 0.014, t = -6.085$, $p < 0.0001$). This is expected given that as discussed in §2.1, ALL had the easiest screens (since the accuracy with ALL, MOST_{SUP} and MOST-OF was very high, we do not see differences due to the difficulty of the screens).

Note that while the semantics of MOST_{SUP} requires comparisons with the two other color sets in the bottom chamber, we see that these comparison procedures have no effect on the accuracy nor on the reaction times. This is compatible with the predictions (as summarized in Table 1) where on the EARLY condition, looks to the target with MOST_{SUP} could benefit from the big set bias in the first two time windows with the rest of the time spent on looking at the other colors; on the LATE condition in the quantifier window there should be more looks to the target and then fewer in the color window than with the other quantifiers. In the next section we will see that the predicted differences are in fact reflected in the eye movements.

2.4 Results: Eye-movements

The pre-processing of the eye-movement data and plotting was carried out using the VWPre package (version 1.2.2, Porretta et al. 2018). The first line in Figures 7–8 shows the plots of the proportion of looks to the target for the EARLY and LATE conditions. The black lines mark the time windows in the audio stimulus adjusted by 200 ms (i.e., 200 ms post the actual onset).[3]

I fitted generalized additive mixed models (GAMMs) using the packages mgcv (version 1.8-31; Wood & Scheipl 2017, Wood 2017) and itsadug (version 2.3; van Rij et al. 2020) to the eye data because a regression line is unable to capture the

[3]200 ms was chosen following Degen & Tanenhaus (2016) because the earliest language mediated fixations are at 200–250 ms after the relevant acoustic landmark that could establish a point of disambiguation (Salverda et al. 2014) The proportion of looks for each interest area has been converted to empirical logits because proportions are inherently bound between 0 and 1 but logits provide a transformation resulting in an unbounded measure suitable for use in the statistical tests (Barr 2008).

nonlinear nature of the time course data as in Figures 7–8. GAMM is a nonlinear regression analysis which in addition to linear effects includes smooth terms as well as random smooths to capture the random effects. Model comparisons involve the full model, with all terms and interactions, and a nested model that excludes the main term and the smooth term corresponding to the predictor and the interactions with these terms (Winter & Wieling 2016, Sóskuthy 2017, Wood 2017). Because in this experiment the predictions are only about the parametric terms (the proportion of looks to the target within a given time window) and not about the differences between the shapes of the curves, significance testing is based on the t-values and I only report those.

The third line in Figures 7–8 shows the model predictions for each of the time windows without random smooths, Preview (−500–0 ms), *Dostałeś* 'You got' (0–1030 ms), the quantifier (1030–2122 ms), *niebieskich/zielonych/czerwonych kulek* 'blue/green/red balls' (2122–3682 ms). The fourth line shows the model predictions including the random smooths that capture the random effects of Subject, Item and Trial. The fifth line summarizes the statistical findings showing which of the contrasts were significant – the unpredicted significant effects are highlighted in grey. The non-highlighted findings match the predictions summarized in Table 1.

In the EARLY condition, Figures 7, there is a main effect of Quantifier in each time window (Preview: $\chi^2 = 10.96, \text{df} = 9, p = 0.009$; 'You got': $\chi^2 = 105.5, \text{df} = 9, p < 0.0001$; the quantifier: $\chi^2 = 55.8, \text{df} = 9, p < 0.0001$; the color window: $\chi^2 = 140.35, \text{df} = 9, p < 0.0001$). Pairwise comparisons reveal the following differences (p-values include the Bonferroni correction for multiple comparisons):

In the Preview window, there are more looks to the target with MOST$_{\text{SUP}}$ than with MOST-OF ($\beta = -0.388, \text{SE} = 0.119, t = -3.268, p = 0.004$) and with SOME than MOST-OF ($\beta = -0.389, \text{SE} = 0.107, t = -3.617, p = 0.001$).

In the 'You got' window, the proportion of looks to the target is higher with MOST$_{\text{SUP}}$ than with SOME ($\beta = -0.653, \text{SE} = 0.102, t = -6.382, p < 0.0001$) and than with MOST-OF ($\beta = -1.259, SE0.111, t = -11.377, p < 0.0001$), as well as with SOME in comparison with MOST-OF ($\beta = -0.604, \text{SE} = 0.101, t = -6.011$, $p < 0.0001$).

In the quantifier window, the trend is reversed and there are fewer looks to the target with MOST$_{\text{SUP}}$ than with SOME ($\beta = 0.637, \text{SE} = 0.115, t = 5.528, p < 0.0001$) and with MOST-OF ($\beta = 0.866, \text{SE} = 0.123, t = 7.02, p < 0.0001$).

In the color adjective plus noun 'balls' window, there are still fewer looks to the target with MOST$_{\text{SUP}}$ than with SOME ($\beta = 0.574, \text{SE} = 0.106, t = 5.436, p < 0.0001$). But now there is no difference between MOST$_{\text{SUP}}$ and MOST-OF. There are now fewer looks to the target with MOST$_{\text{SUP}}$ than with ALL ($\beta = 1.761, \text{SE} =$

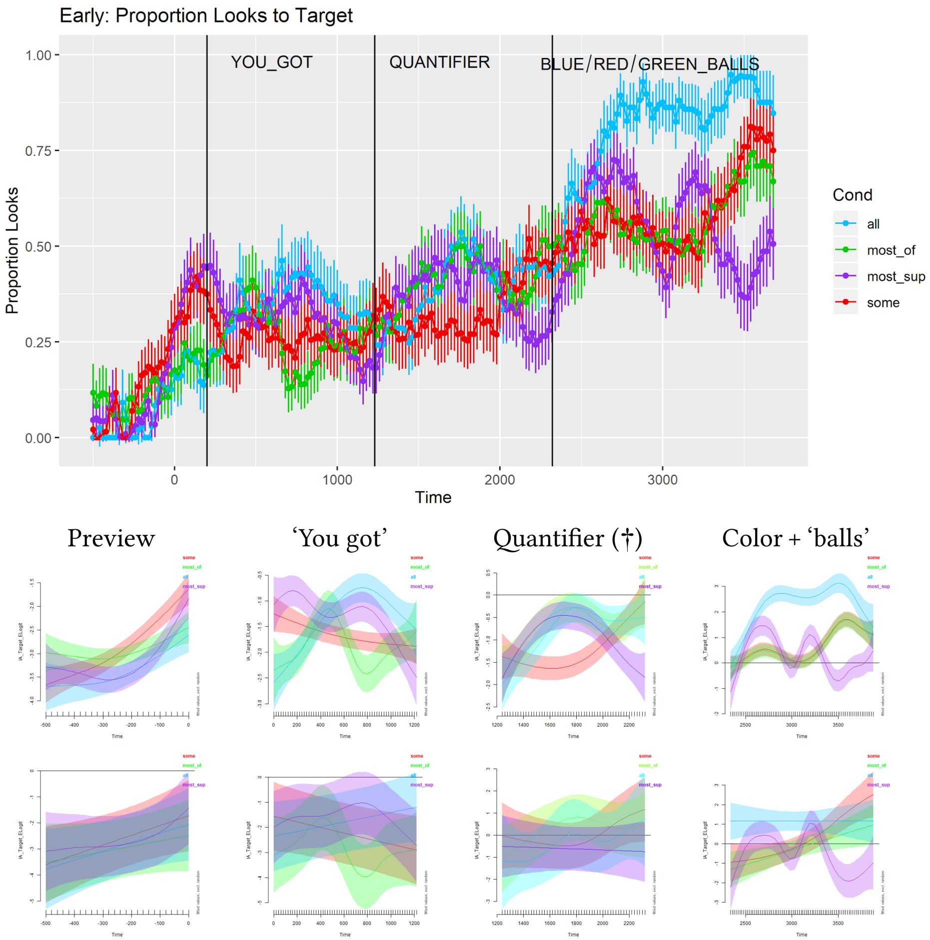
$0.166, t = 10.6, p < 0.0001$). Also SOME has fewer looks to the target than ALL ($\beta = 0.95, \mathrm{SE} = 0.172, t = 5.533, p < 0.0001$), but it has more looks to the target than MOST-OF ($\beta = -0.41, \mathrm{SE} = 0.102, t = -4.003, p = 0.0002$).

Strikingly, there are differences between the quantifiers already during Preview and in the 'You got' window. An exploratory analysis is needed to find out what drove the differences. Perhaps it was some property of the previous trial such as the quantifier, the location of the target (left, center, right) or the numerosities of the sets. Or perhaps this reflects the anticipation of which sentence would best describe the display given that in the LATE condition, there are no differences at Preview, but the differences start at 'You got' such that MOST-OF and SOME get more looks than MOST_{SUP}.[4]

Crucially for the goals of the experiment, in the quantifier window the unexpected trends from the previous windows do not persist. Instead, I find that the predictions have been met: with MOST_{SUP} there were fewer looks to the target than with SOME and MOST-OF. This is the prediction 'MOST_{SUP} < MOST-OF, ALL' in Table 1. (There is no significant difference between MOST_{SUP} and ALL, which was unpredicted, however, SOME is also not significantly different from ALL and it can be seen in the plot with random effects that there is a lot of variation with ALL).

I predicted that the lower proportion of looks to the target with MOST_{SUP} than with MOST-OF should be due the fact that its semantics requires two comparisons (between the target and the two color sets in the lower chamber) while with MOST-OF one comparison is required (between the lower and upper subsets of the partitioned set). The question was whether with SOME the looks to the target would be the same as with MOST_{SUP} suggesting that the processing of the scalar implicature does not happen in the quantifier window. I find that this is not the case: there are more looks to the target with SOME than with MOST_{SUP} and SOME is no different from MOST-OF (and ALL). This result is compatible with the prediction 'SOME-NOT-ALL > MOST_{SUP}' in Table 1, meaning that the scalar implicature, 'some-but-not-all', has already been processed in the quantifier window. I find no support for the alternative, that first the literal meaning of *some*, 'some-and-possibly-all', is processed, 'SOME-POSSIBLY-ALL = MOST_{SUP}' in Table 1.

[4] A reviewer objects to this saying that it is unlikely that the participants would try to guess the upcoming quantifier or were clairvoyant. But my suggestion is that the big-set bias has consequences for the mental representation of the description of the visual scene. See Huettig et al. (2011) for the explanation of the interaction between the visual stimuli and higher order cognitive biases as induced by task goals and language. In the LATE condition, a salient visual cue is absent and the looks do not diverge during Preview. In the EARLY condition, the target set pops out and may bias some mental description of the scene, which is additionally affected by the memory of any salient features of the previous trial.



Preview:

MOST$_{\text{SUP}}$ > MOST-OF

SOME > MOST-OF

MOST$_{\text{SUP}}$ = ALL = SOME

'You got':

MOST$_{\text{SUP}}$ = ALL

MOST$_{\text{SUP}}$ > MOST-OF

MOST$_{\text{SUP}}$ > SOME

SOME > MOST-OF

SOME = ALL

Quantifier (†):

MOST$_{\text{SUP}}$ = ALL

MOST$_{\text{SUP}}$ < MOST-OF

MOST$_{\text{SUP}}$ < SOME

SOME=ALL=MOST-OF

Color + 'balls':

MOST$_{\text{SUP}}$ < ALL

MOST$_{\text{SUP}}$ = MOST-OF

MOST$_{\text{SUP}}$ < SOME

SOME < ALL

SOME > MOST-OF

Figure 7: Results: EARLY condition. (†) marks *disambiguation*.

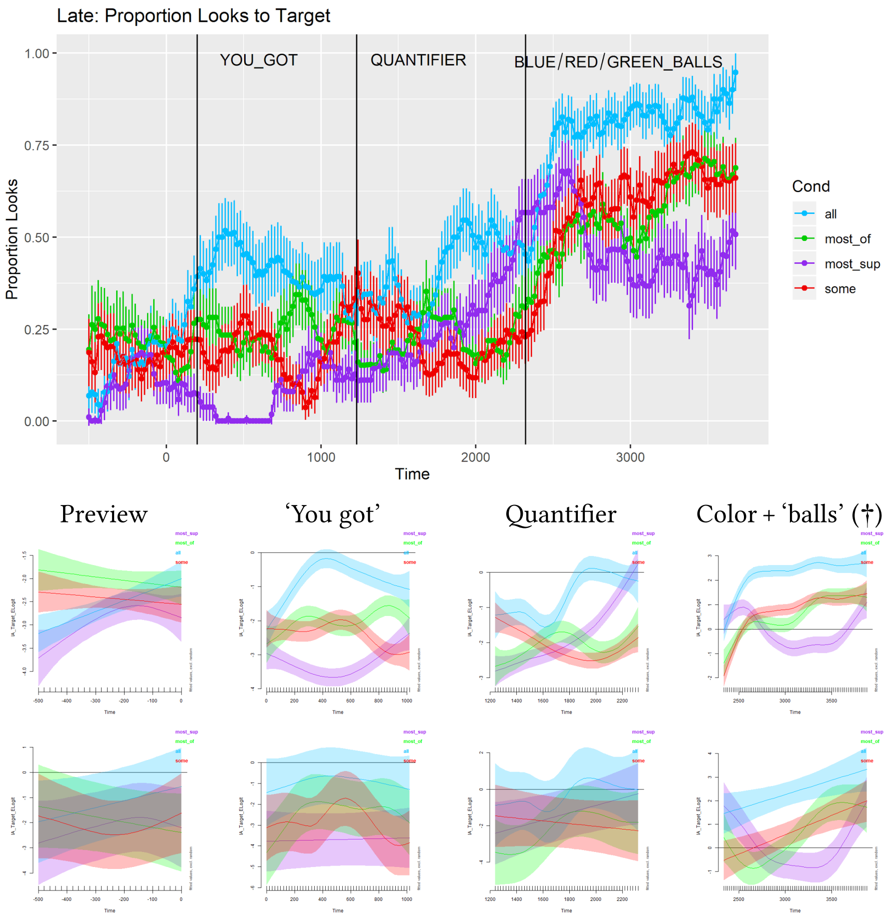
It was also predicted that in the early condition in the quantifier window some-not-all would get more looks to the target than most-of (‘some-not-all > most-of’ in Table 1), but instead we find that some is no different from most-of (‘some = most-of’ in Figure 7) – we do find evidence for this effect but in the next region.

The prediction for the color adjective window was ‘some-not-all > most-of’ if most-of requires more looks between the two partitioned sets than some in order to establish the total set of the balls in the target color for the Subtraction procedure, (2). I hypothesized that the proportion of looks to the target with most-of could be as low as with most$_{\text{sup}}$, which is what we find (‘most$_{\text{sup}}$ = most-of’ in Figure 7). However, looking at the plots we see that the difference between most-of and some is rather small; there are more looks to the target with some at the very beginning and mostly at the end of the region. The trajectory for most-of is quite different than for most$_{\text{sup}}$, where the looks diverge between the target and the distractors. Still, the proportion of looks to the target within the whole region is as low with most-of as with most$_{\text{sup}}$, which is consistent with a higher number of processing steps involved in the Subtraction procedure in contrast with the direct Selection procedure.

The results for the late condition are presented in Figure 8. In the late condition, there are also effects of Quantifier in each time window (Preview: $\chi^2 = 18.4, \text{df} = 9, p = 0.009$; ‘You got’: $\chi^2 = 113.12, \text{df} = 9, p < 0.0001$; the quantifier: $\chi^2 = 112.61, \text{df} = 9, p < 0.0001$; the color window: $\chi^2 = 163.25, \text{df} = 9, p < 0.0001$). In contrast to the early condition, in the Preview window multiple comparisons show no significant differences. In other windows pairwise comparisons reveal the following differences (p-values include the Bonferroni correction for multiple comparisons):

In the ‘You got’ window, in the late condition, most$_{\text{sup}}$ got fewer looks to the target than all the other conditions (some, $\beta = 0.847, \text{SE} = 0.102, t = 8.321, p < 0.001$, all, $\beta = 2.754, \text{SE} = 0.855, t = 3.223, p = 0.005$, most-of, $\beta = 1.274, \text{SE} = 0.108, t = 11.749, p < 0.0001$). some received fewer looks to the target than most-of ($\beta = 0.434, \text{SE} = 0.102, t = 4.258, p < 0.0001$). As in the early condition, this result is unexpected and requires an exploratory analysis.

In the quantifier window, in the late condition, there were more looks to the target with most$_{\text{sup}}$ than with some ($\beta = -0.562, \text{SE} = 0.099, t = -5.688, p < 0.0001$) and than with most-of ($\beta = -0.838, \text{SE} = 0.11, t = -7.606, p < 0.0001$). This is in line with the prediction ‘most$_{\text{sup}}$ > all, most-of, some’ in Table 1 (except that most$_{\text{sup}}$ is not significantly different from all). We also find that there are more looks to the target with some than with most-of ($\beta = -0.276, \text{SE} = 0.103, t = -2.685, p = 0.029$), which fits the prediction ‘some-not-all > most-of’



Preview	'You got'	Quantifier	Color + 'balls' (†)
no differences	MOST$_{SUP}$ < ALL	MOST$_{SUP}$ = ALL	MOST$_{SUP}$ < ALL
	MOST$_{SUP}$ < MOST-OF	MOST$_{SUP}$ > MOST-OF	MOST$_{SUP}$ < MOST-OF
	MOST$_{SUP}$ < SOME	MOST$_{SUP}$ > SOME	MOST$_{SUP}$ < SOME
	SOME = ALL	SOME = ALL	SOME < ALL
	SOME < MOST-OF	SOME > MOST-OF	SOME = MOST-OF

Figure 8: Results: LATE condition. (†) marks *disambiguation*.

in Table 1 and further indicates that with SOME there is no delay in the processing of the scalar implicature.

I predicted that the latter effect would persist in the color window, but instead I find that SOME was no different from MOST-OF. This could be related to the low accuracy with *some*, namely, participants who accepted 'some-not-all' as the description of the display (recall that I analyzed the looks with correct responses only) nevertheless compared the numerosities of the different color sets. However, SOME had significantly fewer looks to the target than ALL ($\beta = 1.676, \mathrm{SE} = 0.276, t = 6.071, p < 0.0001$). This effect, 'SOME-NOT-ALL < ALL' was predicted to occur already in the quantifier window, but we find it later.

In the color window, with MOST_{SUP} there were fewer looks to the target than with the other quantifiers: SOME ($\beta = 0.642, \mathrm{SE} = 0.097, t = 6.641, p < 0.0001$), ALL ($\beta = 2.075, \mathrm{SE} = 0.277, t = 7.505, p < 0.0001$), and MOST-OF ($\beta = 0.46, \mathrm{SE} = 0.101, t = 4.536, p < 0.0001$). This results is exactly as predicted: with MOST_{SUP}, as the color adjective is heard the looks must be directed to the other two color sets in order to make the comparisons to confirm that the target set is indeed the largest.

3 Discussion and conclusions

The results of the current study contribute to the debate about the processing of the scalar implicature of the quantifier *some* during visual verification (Huang & Snedeker 2009, Grodner et al. 2010, Degen & Tanenhaus 2011, Huang & Snedeker 2011, Degen & Tanenhaus 2016) and provide novel predictions for experiments on the processing of *some* in comparison with other quantifiers in languages other than English. I find support for the claim in Spychalska (2009) that the Polish counterpart of *some*, *niektóre*, has a strong implicature – I find that the meaning 'some-not-all' is processed immediately as the disambiguating quantifier is heard. I compared *niektóre* 'some' to *większość* 'most of' and *najwięcej* (the superlative *most*) and *wszystkie* 'all'. In the prior visual world eye-tracking studies, *some* and *all* were compared on the basis of the theory that these two quantifiers form a scale, so when *some* is used instead of the stronger *all*, the inferred meaning is 'some-not-all' (Horn 1972, Levinson 1983, a.o.). However, the results of Degen & Tanenhaus (2011, 2016) showed that whether there is a delay in the processing of the 'some-not-all' meaning depends on whether the experiment contains alternative descriptions using number terms and not just *some* and *all* ('You got some/all/two/three/four/five of the blue gumballs'). When those alternatives are available the processing of the 'some-not-all', implicature is delayed relative to the processing of the meaning of *all*. Without such alternatives, *some* is not delayed relative to *all*.

In the current experiment, adopting the gumball paradigm of Degen & Tanenhaus (2011, 2016) alternative descriptions of the visual scene contained the quantifiers *most of* and the superlative *most* (*most*-SUP) because (i) they allowed for more specific predictions about the time course of the looks to the target than just a comparison with *all*, and (ii) two alternative strategies for *most of* could be tested. Specifically, the semantics of the superlative *most* requires comparisons between the target color set and the two other colors in the lower chambers of the ball machine, which I expected to elicit a distinctive pattern of looks that would serve as the baseline for statistical comparisons (the displays were identical for *some*, *most of* and *most*-SUP, but they had to be different for *all*). As in the study of Degen & Tanenhaus (2016) the quantifiers were compared with two types of displays, EARLY and LATE (Figures 4–5 in §2.2). In the EARLY condition, the quantifier in the stimulus sentence ('You got some/all/most of/the most blue/red/green balls') disambiguated which set of the three sets of balls in the bottom chambers was the target. In the LATE condition, the target was identifiable only when the color adjective was heard.

In the EARLY condition, the results showed no delay for the Polish counterpart of *some* as compared to *all* and *most of*, as well as a higher proportion of looks to the target than with *most*-SUP. I considered two alternatives in the predictions. On the one hand, if the 'some-not-all' meaning was processed early (at the point of hearing the quantifier), the identification of the target set should be just as easy as with *all*, given that the target set was partitioned and as such stood out already during the preview. If, on the other hand, the 'some-possibly-all' meaning was processed first, the looks should be first directed to the unpartitioned sets as with *most*-SUP. The results show support for the first option: the 'some-not-all' meaning is processed early. In the LATE condition, I also find evidence for the 'some-not-all' interpretation in comparison with *all* and *most of*.

The second novel finding concerns the semantics of the majority quantifier *most of*. Pietroski et al. (2009) and Lidz et al. (2011) propose that the verification of sentences like 'You got most of the blue balls' involves a procedure of subtraction (schematized in (2) vs. (3) in §2.2). This procedure requires multiple steps: estimate the superset (the blue balls remaining in the top chamber and the blue balls in the bottom chamber), estimate the target set (the blue balls in the bottom chamber), subtract and compare the result with the target. Subtraction involves more steps than direct comparison of two sets (3 in §2.2), so my hypothesis was that I should find fewer looks to the target in the quantifier and color windows because of the continuing looks to the top blue set in order to establish the total set. Indeed, in the color window there were significantly fewer looks to the target than with both *some* and *all*; the proportion of looks was as low as in the *most*-SUP condition. The low proportion of looks with *most*-SUP can be directly linked to the superlative semantics requiring comparisons with the other color sets. *most of* could be verified by merely comparing the top and bottom numerosities of the partitioned set, but this simple comparison would have elicited a similar proportion of looks to the target as with *some*. The profiles of eye-movements with *some* and *most of* looked similar but the proportion of looks to the target was lower with *most of* than with *some*. In the LATE condition I also predicted fewer looks to the target with *most of* than with *some* in both the quantifier and color windows, and this effect was observed in the quantifier window.

The fact that the pattern of looks with *most of* is compatible with the subtraction procedure and not with the more efficient direct comparison procedure supports the hypothesis in Pietroski et al. (2009) and Lidz et al. (2011) that the mind follows the "instructions" encoded in the lexical representation of quantifier meanings. They argue that lexical semantics interfaces with the cognitive system, which means that lexical meanings require more fine grained distinctions than just truth-conditions. The present experiment showing that with the same display there are distinctive patterns of looks for the three Polish quantifiers *some*, *most of* and *the most* supports the idea that lexical semantics provides direct instructions to visual cognition processes.

Abbreviations

1	first person	PAST	past tense
2	second person	SG	singular
COP	copula	SUP	superlative

Acknowledgments

I would like to thank Piotr Gulgowski for the help with programming in Experiment Builder and for recording participants. Special thanks are due to Yaman Özakın for designing and creating the images of the ball machine. I am very grateful for helpful comments and discussion to Judith Degen, Roumi Pancheva and Maria Spychalska. Thanks are also due to the two anonymous reviewers for very useful comments and suggestions.

This research was funded by the OPUS 5 HS2 grant (DEC-2013/09/B/HS2/02763) from the Polish National Science Center (NCN).

References

Allopenna, Paul D., James S. Magnuson & Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38(4). 419–439. DOI: 10.1006/jmla.1997.2558.

Barr, Dale J. 2008. Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language* 59(4). 457–474. DOI: 10.1016/j.jml.2007.09.002.

Barwise, Jon & Robin Cooper. 1988. Generalized quantifiers and natural language. In Jack Kulas, James H. Fetzer & Terry L. Rankin (eds.), *Philosophy, language, and artificial intelligence: Resources for processing natural language*, 241–301. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-009-2727-8_10.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth & Harald Baayen. 2015. *Parsimonious mixed models.* Ms. Cornell University. https://arxiv.org/abs/1506.04967.

Bates, Douglas, Martin Mächler, Benjamin M Bolker & Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: 10.18637/jss.v067.i01.

Clark, Robin & Murray Grossman. 2007. Number sense and quantifier interpretation. *Topoi* 26(1). 51–62. DOI: 10.1007/s11245-006-9008-2.

Corretge, Ramon. 2020. *Praat Vocal Toolkit.* http://www.praatvocaltoolkit.com.

Degen, Judith & Michael K. Tanenhaus. 2011. Making inferences: The case of scalar implicature processing. *Proceedings of the Annual Meeting of the Cognitive Science Society 33* 33. 3229–3304.

Degen, Judith & Michael K. Tanenhaus. 2016. Availability of alternatives and the processing of scalar implicatures: A visual world eye-tracking study. *Cognitive Science* 40(1). 172–201. DOI: 10.1111/cogs.12227.

Dehaene, Stanislas. 2009. Origins of mathematical intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences* 1156(1). 232–259. DOI: 10.1111/j.1749-6632.2009.04469.x.

Dehaene, Stanislas. 2011. *The number sense: How the mind creates mathematics.* Oxford: Oxford University Press.

Feigenson, Lisa, Stanislas Dehaene & Elizabeth Spelke. 2004. Core systems of number. *Trends in cognitive sciences* 8(7). 307–314. DOI: 10.1016/j.tics.2004.05.002.

Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary & Michael K. Tanenhaus. 2010. "Some," and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116(1). 42–55. DOI: 10.1016/j.cognition.2010.03.014.

Halberda, Justin, Sean F. Sires & Lisa Feigenson. 2006. Multiple spatially overlapping sets can be enumerated in parallel. *Psychological Science* 17(7). 572–576.

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. DOI: 10.1080/15475440801922099.

Horn, Laurence R. 1972. *On the semantic properties of logical operators in English.* Los Angeles, CA: University of California. (Doctoral dissertation). https://linguistics.ucla.edu/images/stories/Horn.1972.pdf.

Huang, Yi Ting & Jesse Snedeker. 2009. Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology* 58(3). 376–415. DOI: 10.1016/j.cogpsych.2008.09.001.

Huang, Yi Ting & Jesse Snedeker. 2011. Logic and conversation revisited: Evidence for a division between semantic and pragmatic content in real-time language comprehension. *Language and Cognitive Processes* 26(8). 1161–1172. DOI: 10.1080/01690965.2010.508641.

Huettig, Falk, Christian NL Olivers & Robert J Hartsuiker. 2011. Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica* 137(2). 138–150. DOI: 10.1016/j.actpsy.2010.07.013.

Hunter, Tim, Jeffrey Lidz, Darko Odic & Alexis Wellwood. 2017. On how verification tasks are related to verification procedures: A reply to Kotek et al. *Natural Language Semantics* 25(2). 91–107. DOI: 10.1007/s11050-016-9130-7.

Knowlton, Tyler, Tim Hunter, Darko Odic, Alexis Wellwood, Justin Halberda, Paul Pietroski & Jeffrey Lidz. 2021. Linguistic meanings as cognitive instructions. *Annals of the New York Academy of Sciences.* DOI: 10.1111/nyas.14618.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. Lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. DOI: 10.18637/jss.v082.i13.

Levinson, Stephen C. 1983. *Pragmatics.* Cambridge: Cambridge University Press. DOI: 10.1017/cbo9780511813313.

Lidz, Jeffrey, Paul Pietroski, Justin Halberda & Tim Hunter. 2011. Interface transparency and the psychosemantics of *most*. *Natural Language Semantics* 19(3). 227–256. DOI: 10.1007/s11050-010-9062-6.

Lindström, Per. 1966. First order predicate logic with generalized quantifiers. *Theoria* 32(3). 186–195. DOI: 10.1111/j.1755-2567.1966.tb00600.x.

McMillan, Corey T., Robin Clark, Peachie Moore, Christian Devita & Murray Grossman. 2005. Neural basis for generalized quantifier comprehension. *Neuropsychologia* 43(12). 1729–1737. DOI: 10.1016/j.neuropsychologia.2005.02.012.

McMillan, Corey T., Robin Clark, Peachie Moore & Murray Grossman. 2006. Quantifier comprehension in corticobasal degeneration. *Brain and Cognition* 62(3). 250–260. DOI: 10.1016/j.bandc.2006.06.005.

Montague, Richard. 1973. The proper treatment of quantification in ordinary english. In Kaarlo Jaakko Juhani Hintikka, Edith A. Moravcsik & Patrick Suppes (eds.), *Approaches to natural language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, 221–242. Dordrecht: Springer Netherlands. DOI: 10.1007/978-94-010-2506-5_10.

Mostowski, Andrzej. 1957. On a generalization of quantifiers. *Fundamenta Mathematicae* 44. 12–36. DOI: 10.4064/fm-44-1-12-36.

Odic, Darko, Paul Pietroski, Tim Hunter, Justin Halberda & Jeffrey Lidz. 2018. Individuals and non-individuals in cognition and semantics: The mass/count distinction and quantity representation. *Glossa* 3(1). DOI: 10.5334/gjgl.409.

Partee, Barbara H. 2011. Formal semantics: Origins, issues, early impact. *The Baltic International Yearbook of Cognition, Logic and Communication* 6(1). 13. DOI: 10.4148/biyclc.v6i0.1580.

Pietroski, Paul, Jeffrey Lidz, Tim Hunter & Justin Halberda. 2009. The meaning of 'most': Semantics, numerosity and psychology. *Mind & Language* 24(5). 554–585. DOI: 10.1111/j.1468-0017.2009.01374.x.

Porretta, Vincent, Aki-Juhani Kyröläinen, Jacolien van Rij & Juhani Järvikivi. 2018. Visual world paradigm data: From preprocessing to nonlinear timecourse analysis. In Ireneusz Czarnowski, Robert J. Howlett & Lakhmi C. Jain (eds.), *Intelligent decision technologies 2017: Proceedings of the 9th KES International Conference on Intelligent Decision Technologies (KES-IDT 2017) – Part II* (Smart Innovation, Systems and Technologies 73), 268–277. Cham: Springer. DOI: 10.1007/978-3-319-59424-8_25.

R Core Team. 2017. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. https://www.r-project.org.

Salverda, Anne Pier, Dave Kleinschmidt & Michael K. Tanenhaus. 2014. Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language* 71(1). 145–163. DOI: 10.1037/e633262013-240.

Sóskuthy, Márton. 2017. *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction*. Ms. University of York. https://arxiv.org/abs/1703.05339v1.

Spychalska, Maria. 2009. *Scalar implicatures and existential import: Experimental study on quantifiers in natural language*. Universiteit van Amsterdam. (MA thesis). https://eprints.illc.uva.nl/id/eprint/817.

Szymanik, Jakub & Marcin Zajenkowski. 2010. Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science* 34(3). 521–532. DOI: 10.1111/j.1551-6709.2009.01078.x.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–1634. DOI: 10.1126/science.7777863.

Tomaszewicz, Barbara. 2011. Verification strategies for two majority quantifiers in Polish. In Ingo Reich (ed.), *Proceedings of Sinn und Bedeutung 15*, 597–612. Saarbrücken: Saarland University Press. https://ojs.ub.uni-konstanz.de/sub/index.php/sub/issue/view/13.

Tomaszewicz, Barbara. 2012. Semantics and visual cognition: The processing of Bulgarian and Polish majority quantifiers. *LSA Annual Meeting Extended Abstracts* 3(0). DOI: 10.3765/exabs.v0i0.603.

Tomaszewicz, Barbara. 2013. Linguistic and visual cognition: Verifying proportional and superlative *most* in Bulgarian and Polish. *Journal of Logic, Language and Information* 22(3). 335–356. DOI: 10.1007/s10849-013-9176-6.

Tomaszewicz, Barbara. 2018. Focus effects on quantifier domains in a visual verification task. *Acta Linguistica Academica* 65(2–3). 417–442. DOI: 10.1556/2062.2018.65.2-3.7.

Troiani, Vanessa, Jonathan E. Peelle, Robin Clark & Murray Grossman. 2009. Is it logical to count on quantifiers? Dissociable neural networks underlying numerical and logical quantifiers. *Neuropsychologia* 47(1). 104–111. DOI: 10.1016/j.neuropsychologia.2008.08.015.

van Rij, Jacolien, Martijn Wieling, R. Harald Baayen & Hedderik van Rijn. 2020. *itsadug: Interpreting Time Series and Autocorrelated Data Using GAMMs.* R package version 2.4. https://cran.r-project.org/web/packages/itsadug/index.html.

Winter, Bodo & Martijn Wieling. 2016. How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. *Journal of Language Evolution* 1(1). 7–18. DOI: 10.1093/jole/lzv003.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R.* London: CRC press. DOI: 10.1201/9781315370279.

Wood, Simon N. & Fabian Scheipl. 2017. *Package gamm4: Generalized additive mixed models using mgcv and lme4.* https://www.rdocumentation.org/packages/gamm4/versions/0.2-6/topics/gamm4.