The effects of quantifier size on the construction of discourse models

Eva Klingvall a,*, Fredrik Heinat b

a Lund University, Centre for Languages and Literature, Box 201, 221 00, Lund, Sweden
b Linnaeus University Department of Languages, Box 451, 351 06, Växjö, Sweden

ARTICLE INFO

Keywords:
Sentence processing
ERP
P600
Semantics
Swedish
Acceptability study

ABSTRACT

Sentences with quantified expressions involve mental representations of sets of individuals for which some property holds (the reference set), as well as of sets for which the property does not hold (the complement set). Both sets can receive discourse focus with negative quantifiers, while the reference set is strongly preferred with positive quantifiers, complement set focus however being possible if contextually motivated. In an offline semantic plausibility study and two online EEG studies, we investigated whether the complement set is an available discourse entity inherently for positive quantifiers, as it is for negative quantifiers. The results show that while the default focus patterns induced by positive and negative quantifiers are robust, both complement and reference set are represented as discourse entities and this is to our knowledge the first study to show that even positive quantifiers make both reference and complement set mentally represented during discourse processing without contextual influence. We also discuss the impact the results from the two ERP studies have on the functional interpretation of two well known ERP effects: the N400 and the P600.

1. Introduction

This paper looks at an essential property of quantifiers, namely that of picking out proportions of different size of some known or implied total quantity. We call this QUANTIFIER SIZE and the question we investigate is what importance this property of quantifiers has in the construction of the mental discourse model (see e.g. von Heusinger & Schumacher, 2019). To interpret a quantified statement such as Most teachers attended the meeting, the relation between ‘teachers attending the meeting’ and ‘teachers not attending the meeting’ has to be considered. At the discourse level, however, usually only one of these groups of ‘teachers’, those attending the meeting, is prominent, salient, or focused, in the discourse (in the sense of Ariel, 1988; Brocher & von Heusinger, 2018; Gundel, Hedberg, & Zacharski, 1993, see also Moxey & Sanford, 1987). Positive and negative quantifiers have been claimed to differ in what discourse entities they make available, and these differences have been argued to be lexical in nature (Kibble, 1997; Nouwen, 2003), or to stem from differences in contextual licensing (Moxey, 2005; Moxey, Sanford, & Dawdydiak, 2001). In this paper we explore this question further, asking what discourse entities quantified expressions make available in the discourse model.

We are particularly interested in the quantified expressions inherent property of size, i.e. to what extent the proportion between ‘teachers attending’ and ‘teachers not attending’ plays any role when the mental discourse model is built up. We investigate whether there are indications that both these groups are mentally represented in the discourse model, even though usually only one is salient

* Corresponding author.
E-mail addresses: Eva.klingvall@englund.lu.se (E. Klingvall), fredrik.heinat@lnu.se (F. Heinat).

https://doi.org/10.1016/j.jneuroling.2022.101066
Received 23 June 2021; Received in revised form 31 January 2022; Accepted 3 February 2022
Available online 24 February 2022
0911-6044/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
enough to be explicit in discourse. We address this question in one off-line judgement study and two online experiments using event-related potentials (ERP). At the same time, the results from the ERP experiments will touch on the question of the functional interpretation of the ERP components N400 and P600. Are they both related to contextual integration of lexical items, or is that a function reserved for the P600 (see discussion and references in Delougou, Brouwer, & Crocker, 2019, 2021)?

One of the functions of quantified expressions (e.g. *some, no, most*, henceforth QEs) is to specify the proportion or number of entities for which some property holds (e.g. Barwise & Cooper, 1981; Keenan & Stavi, 1986). In Fig. 1, we illustrate this using sets: Set A contains ‘teachers’ and Set B contains ‘people attending the meeting’ and the intersection between these contains ‘teachers attending the meeting’. When we use the QE *most*, we say that there are more members in the intersection between the two sets, the striped area in Fig. 1, than in the part of Set A that is not in Set B, the shaded area of Set A. If we instead use the QE *almost no* (almost no teachers attended the meeting), there will be very few members in the intersection compared to members in the rest of Set A. Following Moxey and Sanford (1987), we refer to the intersection as the REFERENCE SET (REFSET) and the part of Set A that is not in Set B as the COMPLEMENT SET (COMPSET).

At the discourse level, *most*, which is a positive QE, and *almost no*, a negative QE, behave quite differently. Both of the sentences in (1) again talk about teachers attending a meeting. When referred back to using an anaphoric pronoun (*they*), negative QEs such as *almost no* make available both the set initially talked about, the REFSET, as in (2a), and the set not mentioned before, the COMPSET, as in (2b). Positive QEs, such as *most*, in contrast, do not readily allow this switch in reference, generally only making the REFSET, (2a), available for reference (e.g. Moxey & Sanford, 1987; Sanford, Moxey, & Paterson, 1996).

| (1) | a. | Almost no teachers attended the meeting … |
|     | b. | Most teachers attended the meeting … |
| (2) | a. | … and that they were present was noteworthy. |
|     | b. | … and that they were absent was noteworthy. |

However, if such a move is contextually motivated, even positive QEs allow for a switch in attentional focus from the REFSET to the COMPSET. Linda Moxey and colleagues have shown that COMPSET focus is closely tied to whether the quantified statement involves a denial of a supposition (e.g. Klingvall & Heinat, 2021; Moxey, 2006; Moxey, Filik, & Paterson, 2009; Moxey et al., 2001; Sanford, Dawydiak, & Moxey, 2007). Most negative QEs inherently give rise to such denials like sentential negation, they come with the inference that the non-negated situation should have applied (Clark, 1976). Thus, from the statement in (1a), we infer that there is a shortfall between teachers expected to attend and actually attending. In such situations, we easily turn our attention to those teachers expected to attend but not attending (the COMPSET), looking for reasons for their absence. Other than expectations, also implicitly or explicitly stated desires can induce shortfall readings (see Ingram & Moxey, 2011; Moxey & Filik, 2010). While positive QEs do not inherently give rise to such shortfall implications, they can be contextually induced. That is, if expectations of a larger proportion are explicitly introduced in the context, even positive QEs, if they pick out a much smaller proportion, can be followed by COMPSET focus (Ingram & Ferguson, 2018; Moxey, 2006; Moxey & Filik, 2010; Moxey et al., 2001, 2009).

If both set readings are possible for both positive and negative QEs (although under much more restricted circumstances for the former), there might also be inherent properties within these categories affecting their focus properties. This is the questions we address in this paper. Existing research has either compared positive and negative QEs (Filik, Leuthold, Moxey, & Sanford, 2011; Heinat & Klingvall, 2019, 2020) or has focused on external factors influencing the attentional focus (e.g. Moxey, 2006; Moxey et al., 2009; Moxey & Filik, 2010; Upadhyay, Houghton, & Klin, 2019). In order to address the question if there are inherent properties in the QEs that influence discourse focus, we need a factor other than context or polarity that can be modulated. One such factor, if not the only, is the proportion between the REFSET and COMPSET (henceforth QE size). Size is known to play a role for our attentional focus in other cognitive domains (Collegio, Nah, Scotti, & Shomstein, 2019), so from that perspective it would not be surprising if it also plays some role for the focus patterns of positive and negative QEs.

The relevance of QE size in language production has been investigated in two offline studies by Sanford et al. (1996). Participants were given one sentence each that included a QE with a ‘near zero’, a ‘mid-to-high’ or a ‘near-to-all’ size of the REFSET (e.g. Few of the parents liked the headmaster’s ideas), and their task was to write a follow-up sentence, beginning with the anaphoric pronoun they. COMPSET continuations were produced to a high degree following all negative QEs, irrespective of QE size, but were not produced at all following positive QEs. Polarity was thus more important than size in these production studies. Notably though, these studies can only tell us that size was not the most important factor when the participants could only write one continuation, but does not tell us whether size still has an effect when a choice between COMPSET and REFSET does not have to be made. To address this issue in the present paper, we move away from production and instead focus on interpretation. More specifically, we ask whether QE size plays a role in acceptability ratings, and also in the online processing of sentences including QEs, focusing on Swedish. Should size play a role in the processing of QEs, there is evidence that the mental discourse model does contain both REFSET and COMPSET, even though only one set is prominent.

---

1. Positive QEs are, by definition, QEs that are monotone increasing, i.e. they make entailments from a subset to a superset: Most teachers wear yellow socks entails Most teachers wear socks. Negative QEs are, by definition, monotone decreasing and make entailments from a superset to a subset: Almost no teachers wear socks entails Almost no teachers wear yellow socks (on QE categories, see e.g. Barwise & Cooper, 1981; Peters & Westerståhl, 2006).

2. Sanford et al. (1996) clearly show that the COMPSET interpretation is different from a MAXSET interpretation (all of Set A), which would be a generalization about teachers as a group, in this case. In
Previous research on QEs in Swedish using offline methods has shown that speaker preferences are even more pronounced than in English: for both positive and negative QEs, only one set is focused at the discourse level – the REFSET for positive QEs, and the COMPSET for negative QEs (Heinat & Klingvall, 2019; Klingvall & Heinat, submitted). In an online processing study, however, there was evidence that both sets indeed compete for focus following negative QEs (Heinat & Klingvall, 2020). Whether that is the case also for positive QEs has not been tested before. As a first step in investigating whether QE size is relevant in processing, we conducted an offline semantic plausibility experiment, Experiment 1, the results of which indicated that online studies of relative set size may reveal something about the discourse models of QEs. In two ERP experiments, we then investigated the online processing of positive QEs, Experiment 2, and negative QEs, Experiment 3.

2. Experiment 1: Semantic plausibility study

From previous research using acceptability ratings, it is clear that positive QEs in Swedish as a group focus the REFSET, while negative QEs focus the COMPSET (Heinat & Klingvall, 2019). In Experiment 1, we wanted to investigate whether such ratings are modulated when the relative set size was manipulated. The broader aim of the study was to learn more about what factors play a role in the discourse model construction of quantified statements.

A number of factors can have an effect on how speakers judge the acceptability of sentences. In addition to syntactic and semantic factors, also processing related factors can play some role for acceptability judgements (Häusler & Juzek, 2017; Juzek, 2015; Schütze, 2011; Schütze & Sproue, 2013). In the present experiment, the syntactic structure was identical in relevant respects in the material. While lexical semantic or pragmatic factors may always play some role, we used a large number of experimental items to reduce the risk of such factors resulting in a consistent pattern in the ratings. Processing, in contrast, was one factor that we predicted would have a particular effect on the ratings. Processing is known to affect acceptability judgements in the way that a condition that requires more processing than some other condition is likely to lead to lower acceptability (Hofmeister, Casasanto, & Sag, 2012; Hofmeister, Culicover, & Winkler, 2015; Hofmeister & Sag, 2010; Miller & Chomsky, 1963). If set size manipulations play a role in processing, we hypothesized that this would also be reflected in the ratings of the sentences. To our knowledge there are no other studies investigating this and we therefore remained open to how set size would affect the ratings. From research in psychology, it has been shown that our mental representations of the size of objects affect our attentional focus (Colleogio et al., 2019). If these effects of size are also relevant in linguistic processing, two scenarios suggest themselves. In the first, a large set will be more salient than a small set, simply because of its size: a large object attracts more attention than a small one. In the second scenario, the number of individual members of the large set will have the opposite effect on the mental representation. In this scenario, a set with fewer members will be more salient because there are fewer members competing for attention. In the following, we use the label LARGE for a QE with a REFSET that is larger than the COMPSET, and the label SMALL for a QE with a REFSET that is smaller than the COMPSET. In Fig. 1, this means that for LARGE QEs (both positive, such as almost all, and negative, such as not exactly all), the REFSET, the striped area, is larger than the COMPSET, the shaded area. For SMALL QEs (both positive, such as a few, and negative, such as almost no), the REFSET, the striped area, is smaller than the COMPSET, the shaded area.

For the first scenario where the large set is seen as more salient, we formulated the following hypotheses for both positive and negative QEs:

A) REFSET reading should be judged as better following LARGE QEs, than following SMALL QEs.
B) COMPSET reading should be judged as better following SMALL QEs, than following LARGE QEs.

For the second scenario, where the number of members in the large set causes difficulties, we formulated hypotheses in the opposite direction:

A) REFSET reading should be judged as better following SMALL QEs, than following LARGE QEs.
B) COMPSET reading should be judged as better following LARGE QEs, than following SMALL QEs.

Fig. 1. Set A: Teachers; Set B: People attending the meeting; Reference set (A∩B, striped area): Teachers attending the meeting; Complement set (A-B, shaded area): Teachers not attending the meeting; White area (B-A): People other than teachers attending the meeting.
2.1. Method & analysis

2.1.1. Participants

Eighty students from Lund University and Linnaeus University, who were all native speakers of Swedish, participated in the study in exchange for a cinema voucher. The analysis is based on 79 participants (1 participant excluded for not complying with the instructions).

2.1.2. Material

In a pre-test, the size of seventeen QEs (eight positive and nine negative) was rated by 596 participants. Out of the seventeen QEs, we selected six that pick out quantities (i.e., REFSET) smaller than 20% of the given total amount, and six QEs that pick out quantities larger than 75%, see Table 1.

Based on these selected QEs, 128 experimental items were constructed. Each QE occurred 42 or 43 times. In addition to a QE in the subject position, the items contained an adjective that was coherent with either a REFSET or a COMPSET reading, as shown below in (3). Out of the 128 item, four received very low ratings in all conditions, indicating that something other than QE size affected the judgements of these items, and they were excluded from the statistical analyses.

The items were manipulated in a 2 × 2 × 2 design: Polarity (POSITIVE (POS) VS NEGATIVE (NEG) QE) × Size of REFSET (LARGE vs SMALL) × Set reading (REFSET (REF) VS COMPSET (COMP) reading, induced by an adjective), as in (3). The sentences were distributed across eight lists in a Latin square design, such that all lists included an equal number of each manipulation but never more than one sentence from the same item. Each list contained 81 filler items, making the total number of sentences each participant read 209.

(3) a. Condition: POS-LARGE-REF
   Det stora flertalet barn somnade direkt och att de var så trotta förvånade föräldrarna.
   'The great majority of children fell asleep immediately and that they were so tired confused the parents.'

b. Condition: POS-LARGE-REF
   Det stora flertalet barn somnade direkt och att de var så uppspelta förvånade föräldrarna.
   'The great majority of children fell asleep immediately and that they were so excited confused the parents.'

c. Condition: POS-SMALL-REF
   Några enstaka barn somnade direkt och att de var så trotta förvånade föräldrarna.
   'A small number of children fell asleep immediately and that they were so tired confused the parents.'

d. Condition: POS-SMALL-REF
   Några enstaka barn somnade direkt och att de var så uppspelta förvånade föräldrarna.
   'A small number of children fell asleep immediately and that they were so excited confused the parents.'

e. Condition: NEG-LARGE-REF
   Inte riktigt alla barn somnade direkt och att de var så trotta förvånade föräldrarna. 'Not quite all children fell asleep immediately and that they were so tired confused the parents.'

f. Condition: NEG-LARGE-REF
   Inte riktigt alla barn somnade direkt och att de var så uppspelta förvånade föräldrarna.
   'Not quite all children fell asleep immediately and that they were so excited confused the parents.'

g. Condition: NEG-SMALL-REF
   Nastan inga barn somnade direkt och att de var så trotta förvånade föräldrarna.
   'Almost no children fell asleep immediately and that they were so tired confused the parents.'

h. Condition: NEG-SMALL-REF
   Nastan inga barn somnade direkt och att de var så uppspelta förvånade föräldrarna.
   'Almost no children fell asleep immediately and that they were so excited confused the parents.'

2.1.3. Method

The experiment was run in computer labs at Lund University and Linnaeus University. The sentences, which were implemented in Google Forms, were presented one at a time on a computer screen. The participants were asked to rate each sentence on a Likert scale from 1 (very unnatural/strange) to 7 (completely natural). Written instructions and practice sentences were given at the beginning of the experiment session.

---

3 The positive QEs are all monotone increasing and give rise to affirmations at the clause level, for instance, in co-ordination they use the pattern ‘och … också’ (‘and … also’) rather than ‘och inte … heller’ (‘and neither’). The negative QEs are all monotone decreasing and give rise to denials at the clause level (for instance, they use the reverse pattern for co-ordination from that of affirmations) (see Klima, 1964; Moxey et al., 2001, among others).

4 The participants were asked to state the number they thought that a particular QE corresponded to out of a given maximum in a specific context. The context was the same in all cases and the participants gave the number for only one QE each. The task was to answer the question: There were 100 students in the auditorium. QE of them had been there before. How many do you think had been there before? (Give your answer in numbers), where QE stands for one of the tested QEs (for details, see Klingvall & Heinat, 2019).

5 Note that the English versions are not literal translations.
2.1.4. Statistical analysis

We followed the recommendations in Juzek (2015) and Häussler and Juzek (2017) and treated the Likert scores as continuous data and used z-scores for the statistical analysis. The analysis was made using R (R Core Team, 2018) and linear mixed effects models (Baayen, Davidson, & Bates, 2008) using the packages lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017) and Emmeans (Lenth, 2018). The fixed effects in the models were Size (LARGE and SMALL QE), Polarity of the QE (POS and NEG) and Set reference (REFSET and COMPSET) and their interaction. Random intercepts and random slopes for participants and items were included as maximally as permitted by the data.

2.1.5. Results

As seen in Fig. 2a, there were differences in all four conditions with positive QEs. For the positive QEs with REFSET continuations (the default condition), ratings for LARGE QEs (REFSET larger than COMPSET) were higher than for SMALL QEs (REFSET smaller than COMPSET). This difference corresponds to the contrast between sentences (3a) and (3c). When positive QEs were followed by COMPSET continuations (the non-default condition), ratings for SMALL QEs (COMPSET > REFSET) were higher than for LARGE QEs. This difference corresponds to the contrast between sentences (3d) and (3b). The statistical analysis (see Table 2) shows that these differences in rating between LARGE and SMALL positive QEs are significant in both the REFSET and the COMPSET conditions.

The negative QEs also showed differences in all conditions, as seen in Fig. 2b. For the negative QEs with COMPSET continuations (the default condition), ratings for SMALL QEs (REFSET smaller than COMPSET) were higher than for LARGE QEs (REFSET larger than COMPSET). This difference corresponds to the contrast between sentences (3h) and (3f). As shown in Table 2, this difference is significant. When negative QEs were followed by REFSET continuations (the non-default condition), ratings for LARGE QEs were slightly higher than for SMALL QEs, but the pairwise comparisons in Table 2 show that the difference is not significant. This difference corresponds to the contrast between sentences (3e) and (3g).

2.2. Discussion of experiment 1

As pointed out above, it has been shown that with contextual manipulations the attentional focus can in some cases fall on the COMPSET even for positive QEs (Moxey, 2006; Moxey & Filik, 2010; Moxey et al., 2001, 2009). This suggests that the COMPSET can be accessible, at least in some cases, for positive QEs. In this experiment, we investigated whether the focus patterns of QEs are affected by the relative size of the REFSET and the COMPSET. We found that set size indeed has an effect on speaker judgements of quantified statements. The lack of previous research regarding the effects of relative set size on attentional focus made this an exploratory experiment. We formulated two opposite hypotheses regarding the direction of the effect of size and the results are compatible with the one on which the largest set is more salient, repeated here:

Ia REFSET reference should be judged as better following LARGE QEs, than following SMALL QEs.
Iia COMPSET reference should be judged as better following SMALL QEs, than following LARGE QEs.

For positive QEs in both set readings, the relative set size affected the ratings. The ratings were significantly higher when the set that the adjective focussed was also the largest set. We interpret these findings as an indication that the COMPSET is taken into account also in the case of positive QEs, even though it is not available for anaphoric reference. More precisely, the reason that SMALL QEs were rated lower than LARGE ones in the REFSET condition could be that the relatively larger COMPSET interferes in this case, and if a large COMPSET makes COMPSET readings better, it also seems that this set is considered in some sense when the sentences are rated.

For negative QEs, there was a significant difference in ratings between SMALL and LARGE QEs only in the default COMPSET condition. As with positive QEs, QEs picking out the larger set (SMALL QEs for COMPSET readings) were rated as significantly better than QEs picking out the smaller set (LARGE QEs for COMPSET readings). Interestingly, this difference was approximately double the size of that for positive QEs. In the non-default REFSET condition, on the other hand, the difference in ratings was in the expected direction but was not large enough to reach significance. Contrary to our expectations thus, size seems to have mattered more in the default condition for negative QEs than it did for positive QEs, but less in the non-default condition. It is not immediately obvious to us why size would play such an important role in the default condition, but not in the non-default condition. One possibility is that the effects are largely driven by the ratings of LARGE negative QEs. These received overall lower ratings than all other QEs, enhancing the difference between LARGE and

---

Table 1
QEs in the Semantic plausibility study (Relative size of REFSET in parentheses).

| Size  | POSITIVE QEs                                    | NEGATIVE QEs                                    |
|-------|------------------------------------------------|------------------------------------------------|
| LARGE | nistan alla ‘almost all’ (88%)                  | inte exakt alla ‘not exactly all’ (88%)         |
|       | i stort sett alla ‘virtually all’ (93%)        | inte precis alla ‘not precisely all’ (85%)     |
|       | det stora flertalet ‘the great majority’ (76%) | inte riktigt alla ‘not quite all’ (87%)         |
| SMALL | några enstaka ‘a small number of’ (7%)         | fäl ‘few’ (14%)                                 |
|       | några ‘some’ (19%)                              | inte många ‘not many’ (18%)                     |
|       | några fäl ‘a few’ (9%)                         | nistan inga ‘almost no’ (7%)                   |
continuations increased for positive QE

Model: pairwise ~ Polarity * (Size | Set), adjust

Pairwise comparisons of QEs (z-scores).

Table 2

| Estimate | Std. Error | df  | t.ratio | p.value |
|----------|------------|-----|---------|---------|
| POLLARGE.REF - POL.SMALL.REF | 0.388      | 0.041 | 748    | 9.571   | <.0001  |
| POLLARGE.COMP - POL.SMALL.COMP | -0.306     | 0.041 | 747    | -7.549  | <.0001  |
| NEGLARGE.COMP - NEGLARGE.SMALL.COMP | -0.659     | 0.041 | 745    | -16.214 | <.0001  |
| NEGLARGE.REF - NEGLARGE.SMALL.REF | 0.099      | 0.041 | 744    | 2.276   | 0.105   |

Model: pairwise ~ Polarity * (Size | Set), adjust = “tukey”.

SMALL in the COMPSET condition and mitigating the differences in the REFSET condition. We think there are some possible explanations for the low ratings of these QEs that are not related to size. All LARGE negative QEs contain the overt negation inte ‘not’, which might in itself lead to lower ratings. In addition, these QEs are stylistically marked to some degree and might therefore differ in use from the other QEs. All of these things are possible reasons for the low ratings (see Schütze, 1996). Given that the LARGE negative QEs were rated lower overall, we think the results for negative QEs should be interpreted with caution.

Thus, QE size has an effect on speaker judgements, although this effect is not strong enough to turn non-default focus readings into preferred ones: COMPSET readings with positive QEs, and REFSET readings with negative QEs were still judged as overall bad. Size manipulations, however, modulate these preferences. This is in fact in line with previous studies investigating the role of expectations and desires on QE focus readings (Moxey, 2006; Moxey et al., 2009; Moxey & Filik, 2010; Upadhyay et al., 2019). In those studies too, the manipulations modulated the patterns but did not reverse them. Moxey and Filik (2010) report, for instance, that the number of COMPSET continuations increased for positive QEs when an explicit desire for ‘all’ was coupled with an actual outcome of ‘a few’, indicating a large shortfall, compared to when the actual outcome was ‘nearly all’, indicating a small shortfall. REFSET continuations were however by far more common in both cases. For the reverse case with negative QEs, a large surplus increased the number of REFSET continuations considerably compared to cases with a small surplus but again did not reverse the pattern.

One possibility is that these differences in rating that we find in Experiment 1 are a reflection of what happens in processing. That is, it might be the case that size manipulations increase the salience of a non-default set to such extent that it causes disruptions in the processing of the sentences. To test this, we designed two ERP experiments where we investigated the processing of LARGE and SMALL positive and negative QEs, Experiments 2 and 3, respectively.

3. Experiment 2: The processing of positive QEs

In order to investigate the effect of relative set size in the processing of positive QEs, we measured the ERP effects on the particular word that explicitly points to either the REFSET or the COMPSET (the adjective trötta/uppspelta ‘tired/interested’), as in (4), illustrating only the LARGE condition, not the SMALL.

(4) a. Condition: POL-LARGE-REF
Det stora flertalet barn somnade direkt och att de var så trötta förvånade föräldrarna.
‘The great majority of children fell asleep immediately and that they were so tired confused the parents.’

b. Condition: POL-LARGE-COMP
Det stora flertalet barn somnade direkt och att de var så uppspelta förvånade föräldrarna.
‘The great majority of children fell asleep immediately and that they were so excited confused the parents.’

With QEs of more equal overall ratings, it is possible that the negative QEs would have patterned more similarly to positive QEs. However, to our knowledge, there are no other large negative QEs in Swedish to test this.
When listeners encounter anaphors, memory representations of possible antecedents in the discourse representation are activated, and as the discourse proceeds, antecedents are integrated into the unfolding discourse (Barkley, Kluender, & Kutas, 2015; van Berkum, Brown, & Hagoort, 1999). The process of activation links a concept with two linguistic forms, the form of the antecedent and the form of the anaphor (for example Mary and she) (see Coopmans & Nieuwland, 2020, and references therein). We assume that the pronoun is interpreted as referring to the default set at this point since there is nothing in the context that would make the non-default set more salient.

Since the REFSET is the default set for anaphoric reference for positive QEs, it should be even easier to establish reference to this set in the condition where the REFSET is the larger set compared to the condition where the COMPSET is the larger set. In the condition with a larger REFSET, both the relative set size and the default focus work together in making this set the most salient one.

Whether the set that is actually targeted in the sentence is coherent with the default REFSET reading is not clear until the adjective (trött/a/uppspent ‘tired/excited’ in 4) is presented. By the time the adjective is processed, the antecedent of the pronoun should be integrated in the discourse model, and as a consequence we expected to find ERP-effects on the adjective in the non-default contexts relative to the default contexts. Two components have been argued to signal difficulties of discourse integration: the N400- and the P600-components.

The N400-component is an enhanced negativity, peaking at 400 ms after word onset. In addition to being indicative of lexicosemantic processing (see Kutas & Federmeier, 2011, for a review), this component is also sensitive to how well a word fits with the discourse model under construction (Filik & Leuthold, 2008; Nieuwland et al., 2019; Nieuwland & van Berkum, 2006b; van Berkum et al., 1999; van Berkum, Zwieterlood, Hagoort, & Brown, 2003). The P600-component is a slightly later ERP component of the opposite polarity. It has been shown to indicate, among other things, a new referent in the discourse (Burkhardt, 2006). Importantly, it also signals a word’s plausibility in a discourse context, irrespective of its semantic relation to the discourse (Kim & Osterhout, 2005), and is an indication of a more demanding inference process of updating the discourse model (Brouwer, Crocker, Venhuizen, & Hoeks, 2017; Brouwer, Fitz, & Hoeks, 2012; Burkhardt, 2007; Heimat & Klingvall, 2020), or of an attempt to revise an initial parse or interpretation (Roll, Horne, & Lindgren, 2007; Van Petten & Luka, 2012).

In the context of positive QEs, it is clear that REFSET reference is always the default reading, but Experiment 1 indicated that the COMPSET can gain salience in the condition where it is larger than the REFSET. In terms of ERP effects, we therefore predicted the following: For adjectives picking out the REFSET, we expected an N400- and/or a P600-effect for SMALL QEs relative to LARGE QEs. In this condition, the adjective is coherent with the default set reading and it should be easier to integrate the adjective when the REFSET is the larger set. In the COMPSET condition, we expected an N400- and/or a P600-effect for LARGE QEs relative to SMALL QEs. In this condition the adjective is not coherent with the default set reading and it should show ease of integration when the COMPSET is the larger set (prediction i).

Finally, given that the REFSET is the default set for positive QEs, irrespective of their size, we also expected a contrast between REFSET and COMPSET for both LARGE and SMALL QEs. More specifically, we expected to see an N400- and/or a P600-effect on the adjectives that are incoherent in the discourse, i.e. adjectives that are congruent with the COMPSET reading, in comparison to adjectives that are congruent with the REFSET reading (prediction ii).

**Predictions:** If both the REFSET and the COMPSET are competing for focus in the processing of statements with positive QEs, there should be.

i N400/P600 effects at the adjective; more specifically, we expected effects for SMALL relative to LARGE in the REFSET condition, and for LARGE relative to SMALL in the COMPSET condition.

ii There should also be a general effect on the adjective in both the size conditions such that the COMPSET should show N400/P600 effects compared to the REFSET.

3.1. Method & material

3.1.1. Participants

Thirty-eight students from Lund University, who were all native speakers of Swedish, participated in the study in exchange for cinema vouchers. They were all right handed and had no diagnosed psychiatric or neurological disorders. One participant completed only half of the experiment due to technical problems and as a consequence the data from that participant have been excluded from the data analysis.

3.1.2. Materials

The materials consisted of 160 experimental items, 124 of these from Experiment 1 and the remaining 36 from Heimat and Klingvall (2020), modified in the same way as for Experiment 1. All of them featured only positive QEs and were manipulated along two dimensions: QE **Size** (LARGE, SMALL) x Set reading (REFSET, COMPSET) of the adjective (Det stora flertalet/Några enstaka barn somnade direkt och att de var så trött/uppspent forvandade foraldrarna. ‘The great majority of/A small number of children fell asleep immediately and that they were so tired/excited surprised the parents.’ See (3a)–(3d) above). The sentences were randomly distributed across four conditions.
lists in a Latin square design, such that each participant saw an equal number of all conditions but only one sentence from each item. There were also 202 unrelated filler sentences included, making the total number of sentences on each list 362.

3.1.3. Procedure

At the very beginning of the session, the participants were given a consent form to complete, also informing them that all data would be anonymized. Once the electrodes had been applied, the participants were placed in front of a computer screen in the lab. On the first slide, instructions were given. These were followed by 10 practice sentences and time for questions. Only after that did the actual experiment start. The stimuli were presented using PsychoPy (Peirce et al., 2019) and each trial began with a cross-hair displayed centrally on the screen for 800 ms, followed by a blank screen for 200 ms after which the sentences were presented word-by-word for 300 ms each, with a 200 ms blank screen interval between them. A message indicating progress appeared on the screen after half of the sentences had been presented, and a second message after two thirds of the sentences. To keep the participants on task, one fourth of the sentences were followed by a yes/no question about the contents of the sentence. No participant was excluded due to low scorings on the questions (correct answers: mean 91%, range 82%–96%). The experiment lasted about 80 min, varying slightly depending on the length of the participants’ pauses and the time they took to answer the questions.

3.1.4. EEG recording and processing

EEG signals were recorded at a sampling rate of 500 Hz using a low impedance system (NeuroScan Easycap), on 30 scalp electrodes in an international 10–20 system, (lateral-left/right: F1/2, F3/4, F7/8, C3/4, T7/8, C3/4, TP7/8, CP3/4, P3/4, P7/8, O1/2, midline: Fz, FCz, Cz, CPz, Pz, Oz), two additional mastoid electrodes, of which the left was used as an online reference, and four eye electrodes (left and right horizontal cantus, and above/below the left eye). The impedance on all scalp electrodes was kept below 5 kΩ for facial electrodes. The EEG was referenced to the average of the left and right mastoid electrodes offline.

The EEG was filtered (0.01–30 Hz band-width filter) and corrected for ocular artefacts using independent component analysis in EELab (Delorme & Makeig, 2004). Data was then segmented into epochs that started 200 ms before critical word onset, and lasted until 1000 ms after adjective onset (trotta/uppspelta, ’tired/excited’ in (3)). All epochs were baseline-corrected and then automatically screened for artefacts (minimal/maximal allowed amplitude = −75/75 μV, and vertical and horizontal eye movements) using ERPLAB (Lopez-Calderon & Luck, 2014). No participant had more than 20% of the epochs removed from analysis (remaining epochs per condition: Adjective – SMALL-COMP (3d): M = 38.4, SD = 1.8, Adjective – LARGE-COMP (3b): M = 38.4, SD = 1.9, Adjective – SMALL-REF (3c): M = 38.5, SD = 1.9, Adjective – LARGE-REF (3a): M = 38.5, SD = 1.5).

3.1.5. Data analysis

Following Luck and Gaspelin (2017), we restricted the analysis to pre-defined regions and time windows relevant for our predictions (see also Nieuwland, 2016). Since we were looking for effects that we suspected were very small, we used the plots from Heinat & Klingvall (2020) and by means of visual inspection refined the region (see Luck & Gaspelin, 2017). The new region was determined before visual inspection and analysis of the data in the present study. The resulting Posterior region consists of CP3, CPz, CP4, P3, P4, O1, OZ and O2 (TP7/8 have been removed from this region compared to Heinat & Klingvall, 2020).

The Posterior region is the region of interest for both N400 effects, 300 ms after critical word onset (see Heinat & Klingvall, 2020; Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010).

For the statistical analyses we performed mixed-effects model analyses (Baayen et al., 2008; Nieuwland, 2016), using the LmerTest package in R (Lenth, 2016).

**Table 3**

| P600     | Estimate | Std. Error | Df   | t value | Pr(>|t|) |
|---------|----------|------------|------|---------|----------|
| LARGE-QUE COMPSET (Int.) | 1.520 | 0.268 | 47.086 | 5.667 | 8.52e-07 ** |
| SMALL-QUE | −0.426 | 0.193 | 272.191 | −2.212 | 0.028 * |
| REFSET | −0.430 | 0.186 | 569.610 | −2.308 | 0.021 * |
| SMALL-QUEREFSET | 0.145 | 0.264 | 635.932 | 0.549 | 0.583 |
| SMALL-QUE REFSET (Int.) | 0.809 | 0.238 | 48.014 | 3.403 | 0.001 ** |
| LARGE-QUE | 0.282 | 0.186 | 257.280 | 1.512 | 0.132 |
| COMPSET | 0.285 | 0.186 | 315.146 | 1.531 | 0.127 |
| LARGE-QUECOMPSET | 0.145 | 0.264 | 632.343 | 0.548 | 0.584 |
| N400     | Estimate | Std. Error | Df   | t value | Pr(>|t|) |
|---------|----------|------------|------|---------|----------|
| LARGE-QUE COMPSET (Int.) | 0.313 | 0.254 | 39.067 | 1.229 | 0.227 |
| SMALL-QUE | −0.369 | 0.189 | 122.553 | −1.946 | 0.054 . |
| REFSET | 0.058 | 0.196 | 74.667 | 0.295 | 0.769 |
| SMALL-QUEREFSET | 0.131 | 0.260 | 398.904 | 0.503 | 0.615 |
| SMALL-QUE REFSET (Int.) | 0.131 | 0.209 | 37.584 | 0.628 | 0.534 |
| LARGE-QUE | 0.239 | 0.178 | 279.312 | 1.338 | 0.182 |
| COMPSET | −0.188 | 0.188 | 172.950 | −1.003 | 0.317 |
| LARGE-QUECOMPSET | 0.130 | 0.260 | 331.042 | 0.502 | 0.616 |

Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.
Model-P600 = lmer(Ampl ~ Size * Set + (1 + Set | Participant) + (1 + Set + Size | Item)).
Model-N400 = lmer(Ampl ~ Size * Set + (1 + Size*Set | Participant) + (1 + Set + Size | Item)).
package for R (Kuznetsova et al., 2017; R Core Team, 2018). The fixed effects in the models were Size (large and small QE) and Set (REFSET and COMPSET) and their interaction. Random intercepts and random slopes for participants and items were included as maximally as permitted by the data. All models reported are the maximal converging models after model comparison.

3.1.6. Results

In the posterior region at the adjective (e.g. ‚trotta/uppspelta ‘tired/excited’), there were significant differences between large QEs and small QEs in the COMPSET condition in the timespan 600–900 ms after critical word onset, see Table 3 and Fig. 3. The top half of Table 3 shows that the contrast between the intercept, that is large QE – COMPSET, and small QE is a significant negativity. This means that in the COMPSET condition, large QEs (COMPSET < REFSET) showed a positivity in relation to small QEs (COMPSET > REFSET). There was also a significant negativity in the condition with large QE – COMPSET (i.e. the intercept) relative to REFSET. In other words, for large QEs, the adjectives that focus the COMPSET showed a positivity in relation to the adjectives that focus the REFSET.

In order to see the results for the other contrasts, the model was re-levelled with small QE – REFSET as intercept, but as seen in the lower part of the top half of Table 3, there were no significant differences.

The results from the 300–500 ms timespan in the posterior region (the N400) are shown in the lower part of Table 3. There was no significant effect in this timespan.

3.2. Discussion of experiment 2

In this experiment, we predicted that if both the REFSET and COMPSET are competing for focus in the processing of positive QEs, there should be an effect of set size at the adjective. We expected effects in both the coherent default REFSET reading and the incoherent COMPSET reading, and indeed found effects in the COMPSET reading. In the COMPSET condition, large QEs gave rise to a P600 effect on the adjective in comparison to small QEs.

A possible explanation for an effect of set size only in the COMPSET condition is that this is the non-default condition. The P600 is often taken to be an indication of an update or reorganization of the discourse model. Burkhardt (2007) sees the P600 as indicating an increased load on discourse memory, visible as a P600 effect, according to Burkhardt. In our study, the adjectives indicating REFSET or COMPSET were always semantically coherent from a lexical point of view (the semantic content of the clauses with the QEs were identical, which is probably why there were no significant N400 effects), but incoherent from a discourse perspective depending on what group had the referential focus. When the incongruent adjective is encountered (the COMPSET condition), the discourse model needs reorganizing and this is when the size of the COMPSET matters. When the COMPSET set is large, it is more salient than when it is small, and thereby also easier to accommodate into the discourse model and less taxing on discourse memory (cf. Burkhardt, 2007).

We also expected to see effects for the non-default COMPSET condition relative to the default REFSET condition for both large and small QEs. For large QEs, there was indeed such a difference in the expected direction in the P600 timespan – COMPSET being more positive than REFSET. We think that these results can be accounted for in line with the discussion in the previous paragraph. In the condition with large QEs and an adjective picking out the REFSET, all factors enhance processing: the focussed set is the largest set, the adjective is coherent with the discourse model. When the adjective picks out the COMPSET, on the other hand, all factors make processing more difficult: Since the COMPSET is not the default set, the discourse model needs restructuring, and since the COMPSET is the smaller set, it is less salient and causes an increased cost for discourse memory.

If all factors conspire to make the updating process more difficult in the condition with large QEs, as described above, the opposite holds for small QEs. As the results show, with small QEs, it is easier to restructure the discourse model in the non-default condition, than it is for large QEs, because for small QEs, the COMPSET is bigger and more salient. The ERPs in the comparison between REFSET-COMPSET reference is therefore not significant in the small condition.

Other results supporting such an account come from a self-paced reading study conducted by Sanford et al. (1996). Sanford and colleagues investigated reading times for sentences very similar to the ones in the present study. A QE (positive or negative) was followed by a sentence containing a phrase targeting either the REFSET or the COMPSET (such as their presence or their absence). As expected, the results from the experiment confirmed that positive QEs favor the REFSET for anaphoric reference and negative QEs the COMPSET. However, what is more interesting from the perspective of our study is that the two positive QEs used by Sanford et al. (1996) differed in size: a few is a small QE, and many is a large QE. While the individual QEs were not compared in the statistics, the data presented indicate that QE size mattered to some degree. In the default REFSET condition there was no real differences in reading times between the QEs, but in the non-default COMPSET condition, the large QE many had markedly longer reading times than the small QE a few. So the same situation obtains in their experiment: a large COMPSET (that is a small QE) facilitates the restructuring of information in the discourse model.

Taken together, we interpret the results from Experiment 2 to mean that QE set size matters in processing only when the information in the discourse model has to be restructured. The larger the COMPSET is, the easier it is to accommodate in the discourse model. We take this to mean that the COMPSET is indeed part of the discourse model, but that the REFSET and the COMPSET are not competing for discourse focus to the extent that there is referential ambiguity.
4. Experiment 3: The processing of negative QEs

From previous studies it is clear that negative QEs differ from positive ones in that both the REFSET and the COMPSET are possible antecedents for anaphoric reference without any contextual manipulations, even though the COMPSET is the favored set (Filik et al., 2011; Heinat & Klingvall, 2019, 2020; Paterson, Sanford, Moxey, & Dawydiak, 1998; Sanford et al., 1996). The rationale for Experiment 3 was to investigate what effects QE size has when both sets are indeed possible antecedents. In Experiment 2 we saw that QE size plays a role primarily in the strongly incoherent condition for positive LARGE QEs. In light of this, our expectations for Experiment 3 were as follows.

As in Experiment 2, we expected to see effects of size on the adjective in the non-default condition, which is the REFSET condition for negative QEs. In this condition, processing should be facilitated if the default COMPSET is small and the REFSET is large. Unlike for positive QEs, we did not expect to see significant overall differences between REFSET and COMPSET readings at the disambiguating adjective within each of the size conditions (i.e. for LARGE and SMALL, respectively), since both sets are available to some extent. No significant differences between REFSET and COMPSET for negative QEs were found in Heinat and Klingvall (2020).

In sum, the prediction for Experiment 3 was the following.

**Prediction:** If size affects anaphoric resolution and if size plays a role for restructuring the discourse model, there should be.

i P600 (/N400) effects at the adjective; more specifically, we expected the adjective following SMALL QEs to show a positivity in the P600 time span (and/or a negativity in the N400 time span) relative to adjectives following LARGE QEs in the REFSET condition.

4.1. Method & material

Procedure, EEG recording, processing and data analysis were identical to what has been described for Experiment 2 above. No participant had more than 20% of the epochs removed from analysis (remaining epochs per condition: Adjective – SMALL-COMP (3h): M = 38.5, SD = 2.0, Adjective – LARGE-COMP (3f): M = 38.6, SD = 1.9, Adjective – SMALL-REF (3g): M = 38.3, SD = 2.2, Adjective – LARGE-REF (3e): M = 38.7, SD = 2.1).

4.1.1. Participants

Thirty-seven students from Lund University, who were all native speakers of Swedish, participated in the study in exchange for cinema vouchers. They were all right handed and had no diagnosed psychiatric or neurological disorders.
4.1.2. Materials

As in Experiment 2, there were 160 experimental items taken from the same sources, manipulated in the same way (Size x Set reference), but all featuring a negative QE (Nåstan inga/Inte riktigt alla barn somnade direkt och att de var så trött/uppsspelt forväntade föräldrarna. ‘Almost no/Not quite all children fell asleep immediately and that they were so tired/excited’ surprised the parents.’ See (3e)-(3h) above). The sentences were distributed across four lists in a Latin square design, such that each participant saw an equal number of all conditions. 222 filler sentences were also included, making the total number of sentences read by each participant 382. No participant was excluded due to low scores on the comprehension questions. The mean for correct answers were 90.3%, and the range 79%–98%.

4.1.3. Results

In the posterior region, there was a marginal effect in the P600 timespan at the adjective (e.g. trött/uppsspelt ‘tired/excited’) for SMALL relative to LARGE in the REFSET condition, as shown in Table 4 and Fig. 4. That is, SMALL QEs were more positive than LARGE ones. The opposite pattern was found in the COMPSET condition; LARGE QEs were more positive than SMALL QEs.

However, this effect did not reach significance. There was no effect in the N400 timespan.

However, as pointed out by Delogu et al. (2019), P600 effects to semantic manipulations vary in their latencies, from 600 ms to above 1000 ms after the onset of the critical word (see Delogu et al., 2019 and references therein). We therefore followed their procedure and divided the P600 time window into three 200 ms time windows with 100 ms onset intervals: 600–800 ms, 700–900 ms and 800–1000 ms, to see where the effect was the most pronounced (Delogu et al., 2019). As we can see in Table 5, there is a significant effect of size in the non-default condition where the adjective focusses the REFSET in the 800–1000 ms timespan.\textsuperscript{9}

4.2. Discussion of experiment 3

Based on the findings in Experiment 2, we expected an effect of size in the non-default REFSET condition on the adjective and indeed found a P600 effect there, although the effect appeared later for negative QEs compared to positive ones. We interpret this effect in line with Experiment 2: when there is an incongruency in the sentence, as with a REFSET reading with negative QEs, the discourse model needs restructuring and this gives rise to a P600 effect. The fact that this effect was later for negative QEs, we think is due to the fact that REFSET focus is not completely ruled out for negative QEs. In the same vein, in their eye-tracking study, Paterson et al. (1998) found only very late effects of set interpretations for negative QEs, compared to positive QEs, and these effects were only present in regions later than the disambiguating region (see also Filik et al. (2011) for differences in the processing of positive and negative QEs measured in ERPs).

5. General discussion

When interpreting quantified statements, we consider the relation between the set of entities for which some relevant property holds, the REFSET, and the set of entities for which the property does not hold, the COMPSET. With negative QEs, such as few and not all, it is clear that both of these sets are also accessible at the discourse level since both can serve as antecedents to anaphoric pronouns (although the COMPSET is the default set). With positive QEs, such as a few and almost all, in contrast, it is much less clear that both sets are accessible (e.g. Moxey & Sanford, 1987; Sanford et al., 1996). It has been shown, however, that with contextual manipulations the attentional focus can in some cases fall on the COMPSET even for positive QEs (Moxey, 2006; Moxey & Filik, 2010; Moxey et al., 2001, 2009). This suggests that the COMPSET is accessible to some degree, after all, for positive QEs. In this paper we set out to investigate if the attentional focus can also be affected by inherent QE properties, other than polarity. In three experiments we looked at this issue, by manipulating the size of the QEs, i.e. investigating the focus properties of QEs whose REFSET is large (and COMPSET is small) and QEs whose REFSET is small (and COMPSET is large). Research in psychology has shown that perceived object size affects attentional focus in general (Collegio et al., 2019). Our experiments suggest that set size also matters for the attentional discourse focus in some particular cases when it comes to language.

Across the three experiments, the same general patterns obtained: large sets showed a benefit over small sets. The extent to which this was manifested in the experiments differed though.

In the semantic plausibility study, Experiment 1, large sets received higher ratings than small sets in both the REFSET and COMPSET conditions for positive QEs, and in the COMPSET condition for negative QEs. This was also the case in the REFSET condition for negative QEs, but not significantly so. These findings are interesting in at least two ways: firstly, the fact that it is larger sets rather than smaller sets that enhance ratings could be interpreted to mean that larger sets receive more attentional focus; secondly, the fact that ratings for positive QEs in the non-default COMPSET condition were affected by set size indicates that properties of the COMPSET are indeed taken into account also for positive QEs when sentences are rated. Although the effects of set size manipulation were not large enough to turn otherwise non-default readings (COMPSET for positive QEs, REFSET for negative QEs) into preferred ones, they did modulate the ratings. This is in fact in line with previous studies investigating the role of expectations and desires on QE focus readings (Ingram & Ferguson, 2018; Moxey, 2006; Moxey et al., 2009; Moxey & Filik, 2018; Upadhyay et al., 2019). In those studies too, the manipulations modulated the patterns but did not reverse them.

\textsuperscript{9} Relevelling shows that there is no significant effect of size in the COMPSET condition.
If both expectations and set size can play a role for the discourse structure of QE, we might ask if these are related, or even the same thing. Recall the ideas argued for by Moxey and colleagues (e.g. Moxey, 2006; Moxey et al., 2009, 2001) that COMPSET focus is more likely to appear when there is a shortfall between an expected (whether implicit or explicit) quantity and a stated quantity. Almost all negative QEs give rise to such shortfall readings by themselves: they all imply that a larger quantity was expected. For these QEs, size will conspire to this effect.

**Table 4**

| 800–1000 ms | Estimate | Std. Error | df  | t value | Pr(> | |t|) |
|-------------|----------|------------|-----|---------|-------|
| LARGE QE COMPSET (Int.) | 0.48     | 0.20       | 71.68 | 2.40    | 0.02  |
| SMALL QE | 0.54     | 0.23       | 114.69 | 2.32    | 0.02  |
| COMPSET | 0.25     | 0.21       | 613.41 | 1.21    | 0.23  |
| SMALL QE COMPSET | −0.65    | 0.30       | 616.00 | −2.18   | 0.03  |

Signif. codes: ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1.

Model = lmer(Ampl ~ Size * Set + (1 + Size | Participant) + (1 + Size | Item)).

**Fig. 4.** Grand average ERPs following adjective onset in the contexts of negative LARGE QEs (red dotted line) and negative SMALL QEs (blue solid line) and focused set (REFSET left, COMPSET right) and topographic maps showing the effect of size (600–900 ms after adjective onset) for the focused sets (REFSET top, COMPSET bottom).

**Table 5**

| P600 | Estimate | Std. Error | Df  | t value | Pr(> | |t|) |
|------|----------|------------|-----|---------|-------|
| LARGE QE COMPSET (Int.) | 1.241    | 0.233      | 55.645 | 5.329   | 1.85e-06 *** |
| SMALL QE | −0.103   | 0.210      | 132.680 | −0.493  | 0.623  |
| REFSET | 0.215    | 0.197      | 308.328 | 1.090   | 0.277  |
| SMALL QE COMPSET | 0.496    | 0.282      | 619.859 | 1.759   | 0.079  |
| SMALL QE REFSET (Int.) | 1.419    | 0.253      | 53.365 | 5.619   | 7.14e-07 *** |
| LARGE QE | −0.393   | 0.210      | 133.016 | −1.874  | 0.063  |
| COMPSET | −0.281   | 0.202      | 312.461 | −1.394  | 0.164  |
| LARGE QE COMPSET | 0.496    | 0.282      | 619.908 | 1.759   | 0.079  |

Signif. codes: ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1.

Model = lmer(Ampl ~ Size * Set + (1 + Size | Participant) + (1 + Size | Item)).
all’) can only have a small shortfall. This is the pattern we see in the plausibility study: COMPSET being rated as better with SMALL QEs (= big COMPSET) than with LARGE ones. For positive QEs, it is more difficult to see how implicit expectations and a potential shortfall would play a role. While the SMALL positive QEs (några ‘some’, några enstaka ‘a small number of’, några få ‘a few’) all have a big COMPSET, it is not clear that they imply that a larger quantity was expected. In the same way, LARGE positive QEs (nåstan alla ‘almost all’, i stort sett alla ‘virtually all’, det stora flertalet ‘the great majority’) do not in themselves seem to imply any expectations of quantities. Effects of size for positive QEs can therefore not as readily be explained in terms of implied expectations and shortfalls. On the other hand, the effects of size show us that the COMPSET is cognitively active also for positive QEs, and this is presumably also why it can receive the attentional focus in those cases where the context helps bring it out.

The online experiment on the processing of positive QEs, Experiment 2, partly confirmed the findings in the offline experiment. In the non-default COMPSET condition, large sets were processed with more ease than small sets, small COMPSETs showing P600 effects relative to large ones at the adjective. In addition, for QEs with a large REFSET (condition LARGE), there was a processing cost at the adjective, when these were coherent with a COMPSET reading rather than the default REFSET reading, again manifested as a P600-effect. For positive QEs, the REFSET is always the default reading, so the discourse model has to be restructured in case the adjective is not coherent with this reading. This restructuring is what causes the P600-effect (cf. Brouwer et al., 2012, 2017; Burkhardt, 2007). Restructuring is less costly when the non-default COMPSET is large than when it is small, as the large set receives more attentional focus in the discourse model. To our knowledge, these are the first results showing that the COMPSET is available in the processing of sentences with positive QEs, without expectations set up by a context.

For negative QEs, Experiment 3, the patterns were similar to those for positive QEs in the nondefault condition, but the effect was slightly later.\footnote{It should be noted that in their investigation of REFSET and COMPSET focus, Filik et al. (2011) found that the ERP effect of reference to the sets was different for positive QEs compared to negative QEs.} It is interesting that the effects of size are found in the non-default condition for both positive and negative QEs. The type of effect found, a P600, indicates that at the point when the adjective that is incoherent with the default reading is encountered, the discourse model has to be restructured. It is in this context that size plays a role: the non-default set receives more attentional focus when it is large than when it is small and therefore causes less disruptions in the processing than small sets do. For positive QEs, the nondefault COMPSET is non-default to a much higher degree than the REFSET is for negative QEs and the effects are thus earlier for these than for negative ones. The default set, for both positive and negative QEs, has so much attentional focus in the discourse model that the size of this set is irrelevant when no restructuring of the model is necessary, i.e. in the default conditions. We think that is why there are no effects of size in these.

The manipulation of size resulted in more differences in offline measures than in online measures. In the offline experiment, size modulated the ratings of both the default and non-default readings, but in the online experiments, there were effects of size only in the non-default conditions. It seems likely that ratings in this case only partly reflect the online processing as measured in the ERP experiments. They may also reflect effects that appear much later in the processing, at stages later than we have examined.

Finally, the results in the two ERP experiments are interesting from a more general neurolinguistic perspective. They have direct bearing on the ongoing debate about the functional interpretation of the two ERP components we investigated, the N400 and the P600. The fact that the lexical manipulation of the adjective gives rise to only a P600 effect, but no N400 effect for both positive and negative QEs in our experiments can be accounted for in a view that sees the N400 as an indicator of lexical retrieval, but not semantic/discourse integration, and the P600 as an indicator of semantic/discourse integration (see e.g. Brouwer et al., 2017, 2012; Delogu et al., 2019, and references therein). In particular in the case of positive QEs, the results from Exp. 1 show that the adjectives focussing the COMPSET are rated very low, i.e. they are not given by the context in comparison to the adjectives focussing the REFSET. If the N400 indicates the predictability and integration of word meaning into the discourse context, we would expect to see an N400 effect on the adjective in the non-default conditions, COMPSET for positive QEs and REFSET for negative QEs. However, we do not find an N400 effect. This result is fully compatible with an approach that sees the N400 as indicator of access/retrieval, but not contextual integration (Brouwer et al., 2012; Kutas & Federmeier, 2011; Lau, Phillips, & Poeppel, 2008; van Berkum, 2009). It should be pointed out, though, that the experiments were not set up to address this question specifically and we will refrain from drawing too far reaching conclusions on a null result regarding the N400, and just note that the results are compatible with such an approach, but that the experiments lack a manipulation directly addressing the question of lexical prediction and discourse integration.

To conclude, this paper confirms the picture from previous studies that negative QEs have the possibility of focusing either the COMPSET or the REFSET. The results from Experiments 1 and 3 show that set size does play a role for the processing and acceptability of the sentences with negative QEs. In addition, this study shows that positive QEs focus the REFSET very early on in the processing, making anomalies immediately detectable. More important, however, is the new finding that when the anomaly is detected, the results from both Experiments 1 and 2 indicate that the COMPSET is part of the discourse model even for positive QEs. The larger the COMPSET is, the easier the information in the discourse model can be updated. Crucially, the effects we see from the COMPSET with positive QEs are not induced by contextual manipulations such as an explicit or implied shortfall, but by an inherent property of QEs, size. It may be the case that it in fact is the availability of the COMPSET in the discourse model that makes contextual manipulation of positive QEs possible at all. In order to gain a more complete picture of how quantified expressions are integrated in discourse models, we think this interaction between inherent qualities of QEs, on the one hand, and context and the expectations it induces in language users, on the other hand, must be explored in more detail.
Author statement

Both authors contributed equally to the paper.

Ethical approval

The study was performed in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration. The study does not include any minors and does not need an approval from the Ethical Board according to current legislation.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was funded by a grant from The Swedish Research Council, project id 2018-01074 (to FH), a grant from Stiftelsen Olle Engkvist Byggmästare, project id 191-437 (to FH), and a grant from Erik Wallanders fond (to EK). We thank the anonymous reviewers, our participants and Henrik Garde for help with Matlab and Psychopy. We gratefully acknowledge Lund University Humanities Lab. Both authors contributed equally to this research.

References

Ariel, M. (1988). Referring and accessibility. Journal of Linguistics, 24(1), 65–87.
Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. Journal of Memory and Language, 5 (4), 390–412.
Barkley, C., Kluender, R., & Kutas, M. (2015). Referential processing in the human brain: An event-related potential (ERP) study. Brain Research, 1629(5), 143–159.
Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language. Linguistics and Philosophy, 4, 159–214.
van Berkum, J. J. (2009). The neuropragmatics of ‘simple’ utterance comprehension: An ERP review. In K. Yatsuhashi, & U. Sauerland (Eds.), Semantics and pragmatics: From experiment to theory (pp. 276–316). Basingstoke: Palgrave Macmillan.
van Berkum, J. J., Brown, C. M., & Hagoort, P. (1999). Early referential context effects in sentence processing: Evidence from event-related brain potentials. Journal of Memory and Language, 41(2), 147–182.
van Berkum, J. J., Koomen, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. Brain Research, 1146, 158–171. https://doi.org/10.1016/j.brainres.2006.06.091
van Berkum, J. J., Zwitserlood, P., Bastaansen, M., Brown, C., & Hagoort, P. (2004). So who’s “he” anyway? Differential ERP and ERSP effects of referential success, ambiguity and failure during spoken language comprehension. In Annual meeting of the cognitive neuroscience society (cns-2004). San Francisco, April 18–20.
van Berkum, J. J., Zwitserlood, P., Hagoort, P., & Brown, C. M. (2003). When and how do listeners relate a sentence to the wider discourse? Evidence from the N400 effect. Cognitive Brain Research, 17(3), 701–718. https://doi.org/10.1016/S0926-4114(03)00196-4
Brocher, A., & von Heusinger, K. (2018). A dual-process activation model: Processing definiteness and information status. Glossa, 3, 1–34.
Brouwer, H., Crocker, M. W., Venuhuizen, N. J., & Hoeks, J. C. J. (2017). A neurocomputational model of the N400 and the P600 in language processing. Cognitive Science, 41(S6), 1318–1352. https://doi.org/10.1111/cogs.12461
Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. Brain Research, 1446, 127–143.
Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. Brain and Language, 98(2), 159–168. https://doi.org/10.1016/j.bandl.2006.04.005
Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. Neuereport, 18(17), 1851–1854.
Clark, H. H. (1976). Semantics and comprehension. The Hague: Mouton.
Collegio, A. J., Nah, J. C., Scotti, P. S., & Shomstein, S. (2019). Attention scales according to inferred real-world object size. Nature Human Behaviour, 3(1), 40–47.
Coopmans, C. W., & Nieuwland, M. S. (2020). Dissociating activation and integration of discourse referents: Evidence from ERPs and oscillations. Cortex, 126, 83–106.
Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. Brain and Cognition, 135, 103569. https://doi.org/10.1016/j.bandc.2019.05.007
Delogu, F., Brouwer, H., & Crocker, M. W. (2021). When components collide: Spatiotemporal overlap of the N400 and P600 in language comprehension. Brain Research, 1766, 147514. https://doi.org/10.1016/j.brainres.2021.147514
Delorme, A., & Makeig, S. (2004). Eeglab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9–21.
Filk, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: Evidence from the N400. Psychophysiology, 45(4), 554–558. https://doi.org/10.1111/j.1469-8986.2008.00566.x
Filk, R., Leuthold, H., Moxey, L. M., & Sanford, A. J. (2011). Anaphoric reference to quantified antecedents: An event-related brain potential study. Neuropsychologia, 49(13), 3786–3794.
Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. Language, 69(2), 274–307.
Hausser, J., & Juez, T. (2017). Hot topics surrounding acceptability judgement tasks. In S. Featherston, R. Hornig, R. Steinberg, B. Umbreit, & J. Wallis (Eds.), Proceedings of linguistic evidence 2016 - empirical, theoretical, and computational perspectives (pp. 1–23). University of Tübingen, online publication system.
Heinat, F., & Klingvall, E. (2019). Discourse prominence: Definition and application. Journal of Pragmatics, 154, 117–127.
Hofmeister, P., Casasanto, L. S., & Sag, I. A. (2012). How do individual cognitive differences relate to acceptability judgments?: A reply to Sprouse, wagers, and Phillips. Language, 88(2), 390–400.
Hofmeister, P., Culicover, P., & Winkler, S. (2015). Effects of processing on the acceptability of “frozen” extraposed constituents. Syntax, 18(4), 464–483.
Hofmeister, P., & Sag, I. A. (2010). Cognitive constraints and island effects. Language, 86(2), 366–415.
