J Log Lang Inf (2014) 23:53–81
DOI 10.1007/s10849-014-9197-9

On the Identification of Quantifiers' Witness Sets: A Study of Multi-quantifier Sentences

Livio Robaldo · Jakub Szymanik · Ben Meijering

Published online: 4 April 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Natural language sentences that talk about two or more sets of entities can be assigned various readings. The ones in which the sets are independent of one another are particularly challenging from the formal point of view. In this paper we will call them 'Independent Set (IS) readings'. Cumulative and collective readings are paradigmatic examples of IS readings. Most approaches aiming at representing the meaning of IS readings implement some kind of *maximality conditions* on the witness sets involved. Two kinds of maximization have been proposed in the literature: 'Local' and 'Global' maximization. In this paper, we present an online questionnaire whose results appear to support Local maximization. The latter seems to capture the proper interplay between the semantics and the pragmatics of multi-quantifier sentences, provided that witness sets are selected on pragmatic grounds.

1 Prologue

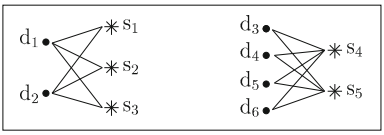
Before starting, we would like that the reader tries to evaluate the truth/falsity of some sentences. The statements are about figures showing dots connected to stars. In the next figure, is it true that "Less than half of the dots are connected with exactly three stars"?

L. Robaldo (✉)
Department of Computer Science, University of Turin, Turin, Italy
e-mail: robaldo@di.unito.it

J. Szymanik
Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherland
e-mail: jakub.szymanik@gmail.com

B. Meijering
Institute of Artificial Intelligence, University of Groningen, Groningen, The Netherland
e-mail: b.meijering@ai.rug.nl

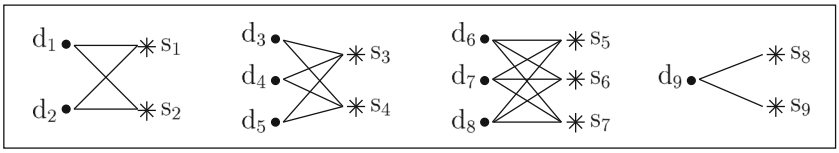
(1)



We do think that the answer is "Yes". The same answer has been given by several friends or colleagues that were asked to judge the example. In fact, the figure does contain two dots d_1 and d_2, which are less than half of all the dots in the figure, and they are both connected with three same stars s_1, s_2, and s_3.

Now, is it true in (2) that "Few dots are connected with few stars"?

(2)



It is somehow harder to provide an answer to this second question. At first sight, it seems the sentence is false, or at least 'strange': probably no English speaker would ever utter that sentence in that context, whatever he wants to describe.

In the literature, most approaches aiming at formally defining the truth values of sentences as such, e.g., (Schein 1993; Landman 2000; Brasoveanu 2012) among others, represent the two sentences under examination via formulae that are false in contexts (1) and (2) respectively. On the other hand, other approaches, e.g., (Sher 1997; Robaldo 2010a), represent the two sentences via formulae that are true in the two contexts respectively.

It seems then that none of the mentioned approaches is completely satisfactory. In our view, the reason is that none of them take enough into account pragmatic preferences involved in witness sets' selection. This paper presents an experimental analysis of some Italian sentences whose results seem to show that such pragmatic factors are indeed at stake. We conclude by advocating a formal solution able to properly include them in the formulae.

2 Introduction

This paper is about Independent Set (IS) readings, a.k.a. Scopeless readings. The well-known cumulative and collective readings (Scha 1981; Landman 2000; Beck and Sauerland 2000; Kratzer 2007; Kontinen and Szymanik 2008; Szymanik 2010) are archetypal examples of IS readings. This paper, following approaches of (van der Does 1993; Schwarzschild 1996; Kratzer 2007; Robaldo 2011), assumes there is a single class of IS readings, known as "cover readings", of which the cumulative and collective readings are merely special cases. A general example of cover reading is shown in (3).

(3) Exactly three children ate exactly five pizzas.

A possible meaning of (3) is that there are three children that, together, as a team, ate a total of five pizzas. In other words, five pizzas is the cumulation of the pizzas eaten by the children. Each pizza was not necessarily eaten as a whole by a child. Rather, it is

likely that it has been cut into slices and shared among two or three children. In formal terms, such an interpretation is satisfied, for instance, by the following extension of ate':

(4) $\|ate'\|^M \equiv \{\langle c_1 \oplus c_2 \oplus c_3, p_1 \oplus p_2\rangle, \langle c_2 \oplus c_3, p_3 \oplus p_4\rangle, \langle c_3, p_5\rangle\}$

'$\oplus$' is the standard sum operator originally introduced by Link (1983). Assuming that a and b are two atomic individuals of the domain, '$a \oplus b$' denotes another individual referring to the set $\{a, b\}$, i.e., the team of the two individuals a and b. Therefore, in (4), children c_1, c_2, and c_3, as a team, share pizzas p_1 and p_2, the group of the two children c_2 and c_3 share p_3 and p_4, and c_3 ate pizza p_5 on his own.

IS readings have always been problematic for theories of quantifier scope, among them Quantifier-Raising (Heim 1982; May 1985; Diesing 1992), Quasi Logical Form (Alshawi 1992), Quantifying-in (Montague 1974), the Cooper Storage (Cooper 1983; Keller 1988), and Discourse Representation Theory (Kamp and Reyle 1993). In fact, such theories by and large neglect them.

For devising a logical formalism able to represent IS readings, it is necessary to depart from standard quantifier embedding operation, i.e., from logical forms that represent different scopings by nesting quantifiers within the scope of other quantifiers (cf. Robaldo 2010b). On the contrary, the formula must include suitable 2-order variables that explicitly denote the witness sets (Sher 1997; Landman 2000; Robaldo 2011; Brasoveanu 2012). The different scopes can be represented by setting such variables to be functionally dependent. In practice, the logic must implement a kind of set-Skolemization among witness sets. In this paper, however, we will only consider IS readings of simple sentences subject-verb-object. Exactly two witness sets are involved for each sentence, and no functional dependency among them is established.

Set-Skolemization does not suffice to properly represent the semantics of *any* IS reading, with *any* quantifier. The formulae have to be further restricted via special clauses, termed *maximality conditions*, when the NPs involve downward[1] monotone (M↓) or non-monotone (non-M) quantifiers (van Benthem 1986; Sher 1990; Spaan 1996; Landman 1998; Steedman 2012). To see why, consider (5.a–c), respectively involving an M↑ quantifier (*At least two*), an M↓ quantifier (*At most two*), and a non-M quantifier (*Exactly two*).

(5) a. At least two men walk.
b. At most two men walk.
c. Exactly two men walk.

It seems that (5.a–c) can be represented via (6.a–c), where M is a second order Skolem constant denoting the set of (at least/at most/exactly) two men.

(6) a. $\exists_M[\,|M| \geq 2 \wedge \forall_x[\,(x \in M) \rightarrow (man'(x) \wedge walk'(x))]\,]$
b. $\exists_M[\,|M| \leq 2 \wedge \forall_x[\,(x \in M) \rightarrow (man'(x) \wedge walk'(x))]\,]$
c. $\exists_M[\,|M| = 2 \wedge \forall_x[\,(x \in M) \rightarrow (man'(x) \wedge walk'(x))]\,]$

Nevertheless, only (6.a) captures the truth values of the corresponding sentence. This becomes clear when one considers a model in which three men walk. In such a model, (5.a) is true, while (5.b) and (5.c) are false. Conversely, all the three formulae in (6)

[1] See Peters and Westerståhl (2006) for a survey on possible monotonicities featured by GQs and Szymanik and Zajenkowski (2013) for a recent computational account.

evaluate to true, in that all of them allow to choose a Skolem constant M denoting a set of two walking men. Therefore, we cannot allow a free choice of the constants occurring in the formulae, but this has to be further constrained. In particular, they have to denote the *maximal* set of individuals satisfying the predicates, e.g., the maximal set of walking men, in the examples in (6). This may be achieved by changing (6.b) and (6.c) to (7.b) and (7.c) respectively. Note that, for the sake of uniformity, the maximality condition can be inserted in (6.a) as well, thus obtaining (7.a). The truth values are not affected.

(7) a. $\exists_M[\,|M| \geq 2 \wedge \forall_x[\,(x \in M) \rightarrow (man'(x) \wedge walk'(x))]\wedge$
$\neg\exists_{\boldsymbol{M'}}[\,\boldsymbol{M \subsetneq M'} \wedge \forall_{\boldsymbol{y}}[\,(\boldsymbol{y \in M'}) \rightarrow (\boldsymbol{man'(y)} \wedge \boldsymbol{walk'(y)})]]]$

b. $\exists_M[\,|M| \leq 2 \wedge \forall_x[\,(x \in M) \rightarrow (man'(x) \wedge walk'(x))]\wedge$
$\neg\exists_{\boldsymbol{M'}}[\,\boldsymbol{M \subsetneq M'} \wedge \forall_{\boldsymbol{y}}[\,(\boldsymbol{y \in M'}) \rightarrow (\boldsymbol{man'(y)} \wedge \boldsymbol{walk'(y)})]]]$

c. $\exists_M[\,|M| = 2 \wedge \forall_x[\,(x \in M) \rightarrow (man'(x) \wedge walk'(x))]\wedge$
$\neg\exists_{\boldsymbol{M'}}[\,\boldsymbol{M \subsetneq M'} \wedge \forall_{\boldsymbol{y}}[\,(\boldsymbol{y \in M'}) \rightarrow (\boldsymbol{man'(y)} \wedge \boldsymbol{walk'(y)})]]]$

The clauses in boldface are maximality conditions asserting the nonexistence of a superset of M whose elements also satisfy man' and $walk'$. (7.a–b) are true iff M denotes the plural individual corresponding to the set of *all* walking men, and this contains exactly two and at most two individuals respectively. In a model including three walking men, (7.a) and (7.b) correctly turn out to be false.

However, as explained below, it is still unclear in the literature how maximality conditions should be asserted when two or more sets of individuals are involved, i.e., for the proper representation of IS readings like (3). Till now, two non-equivalent kinds of maximality conditions have been proposed. This paper discusses an online questionnaire we conducted to investigate how humans interpret IS readings in Italian. The results seem to show that pragmatics plays a crucial role in the interpretation of the IS readings. Currently, only one of the two kinds of maximality conditions is able to properly account for such pragmatic factors.

However, our conclusions do not merely state that the latter is necessarily preferred. Even though our empirical work does not refer to any concrete semantic formalism, we do advocate on theoretical grounds a different *conception* of maximality conditions. They must not be seen as mere "constraints" that need to be satisfied in order to state that a certain formula is true or false in a certain context. Rather, they must be seen as part of the asserted knowledge, needed to make inferences about the sets of entities at stake.

3 Maximality Conditions on Multi-quantifier Sentences

To our knowledge, the first who proposed a relevant generalization of the formulae in (7) for dealing with two or more sets of individuals has been Sher (1990, 1997). However, the framework of (Sher 1997) is able to represent only a very particular class of IS readings, known as Branching Quantifier readings, that are very rare in natural language (Gierasimczuk and Szymanik 2009). Such readings require the cartesian product of the witness sets to be included in the main predicate's extension. Therefore, with respect to two sets S_1 and S_2, each individual in S_1 must be related, via the semantic relation denoted by the main predicate, with each individual in S_2 (and vice versa).

Sentence (3) does not have such a reading, because (parts of) pizzas cannot be eaten by more than one child. On the other hand, the Branching Quantifier reading is available for the examples shown in the Prologue. More generally, Branching Quantifier readings only involve semantic relations, like "be connected", that allow mutual sharing of the items in the object among those in the subject.

In Sher's theory, sentence (8) has a Branching Quantifier reading that is represented via the formula (9). The formula states that there are two sets $\|P_1\|^M$ and $\|P_2\|^M$, respectively including two dots and three stars, such that their cartesian product $\|P_1\|^M \times \|P_2\|^M$ is included in the extension of *connect′*. Clauses in boldface are maximality conditions: there is not a super-cartesian product included in *connect′*'s extension.

(8) Exactly two dots are connected with exactly three stars.

(9) $\exists P_1 P_2[\, 2!_x\, (\text{dot}'(x),\, P_1(x))] \wedge 3!_y\, (\text{star}'(y),\, P_2(y)) \wedge$
$\forall_{xy}[(P_1(x) \wedge P_2(y)) \rightarrow \text{connect}'(x, y)] \wedge$
$\forall_{\boldsymbol{P_1' P_2'}}[\,(\forall_{xy}[(\boldsymbol{P_1}(x) \wedge \boldsymbol{P_2}(y)) \rightarrow (\boldsymbol{P_1'}(x) \wedge \boldsymbol{P_2'}(y))] \wedge$
$\forall_{xy}[(\boldsymbol{P_1'}(x) \wedge \boldsymbol{P_2'}(y)) \rightarrow \textbf{connect'}(x, y)]\,) \rightarrow$
$\forall_{xy}[(\boldsymbol{P_1'}(x) \wedge \boldsymbol{P_2'}(y)) \rightarrow (\boldsymbol{P_1}(x) \wedge \boldsymbol{P_2}(y))]]]$

The scenario shown above in (1) satisfies (9), because it includes two sets $\|P_1\|^M \equiv \{d_1, d_2\}$ and $\|P_2\|^M \equiv \{s_1, s_2, s_3\}$ such that their cartesian product is included in the main predicate's extension. In other words, it holds:

(10) $$\|P_1\|^M \times \|P_2\|^M \equiv$$
$$\{\langle d_1, s_1\rangle, \langle d_1, s_2\rangle, \langle d_1, s_3\rangle, \langle d_2, s_1\rangle, \langle d_2, s_2\rangle, \langle d_2, s_3\rangle\} \subseteq \|connect'\|^M$$

And the model does not include a super-cartesian product of $\|P_1\|^M \times \|P_2\|^M$ still included in $\|connect'\|^M$. In other words, it is not possible to add a dot in $\|P_1\|^M$ or a star in $\|P_2\|^M$ such that each dot in the former would be still connected with each star in the latter.

Despite the limited linguistic coverage, the insights of Sher's approach were rather promising, and worth to be generalized [cf. with the strategy of bounded composition suggested by Dalrymple et al. (1998) and the determiner fitting operator of Winter (2001)]. This was done by Robaldo (2010a, 2011). The latter may be also considered as a generalization of the approach of (Scha 1981), focused on the notion of maximality of the witness sets. The formula proposed by Robaldo (2011) for representing the meaning of sentence (3) is shown in (11).

(11) $3!_x\, (\text{child'}(x),\, P_1(x)) \wedge 5!_y\, (\text{pizza'}(y),\, P_2(y)) \wedge$
$\forall_x[P_1(x) \rightarrow \text{child'}(x)] \wedge \forall_y[P_2(y) \rightarrow \text{pizza'}(y)] \wedge$
$Cover(C, P_1, P_2) \wedge \forall_{xy}[C(x, y) \rightarrow \text{ate'}(x, y)] \wedge$
$\forall_{P_1'}[(\forall_x[P_1(x) \rightarrow P_1'(x)] \wedge \forall_x[P_1'(x) \rightarrow \text{child'}(x)] \wedge$
$\exists_{C'}[Cover(C', P_1', P_2) \wedge \forall_{xy}[C(x, y) \rightarrow C'(x, y)] \wedge \forall_{xy}[C'(x, y) \rightarrow \text{ate'}(x, y)]]) \rightarrow$
$\forall_x[P_1'(x) \rightarrow P_1(x)]] \wedge$
$\forall_{P_2'}[(\forall_y[P_2(y) \rightarrow P_2'(y)] \wedge \forall_y[P_2'(y) \rightarrow \text{pizza'}(y)] \wedge$
$\exists_{C'}[Cover(C', P_1, P_2') \wedge \forall_{xy}[C(x, y) \rightarrow C'(x, y)] \wedge \forall_{xy}[C'(x, y) \rightarrow \text{ate'}(x, y)]]) \rightarrow$
$\forall_y[P_2'(y) \rightarrow P_2(y)]]$

Cover is a meta-predicate stating that every individual in $\|P_1\|^{M,g}$ and $\|P_2\|^{M,g}$ occurs somewhere in the variable $\|C\|^{M,g}$, and that the latter is made up only from individuals in $\|P_1\|^{M,g}$ and $\|P_2\|^{M,g}$. In practice, $\|C\|^{M,g}$ describes how the individuals in the witness sets $\|P_1\|^{M,g}$ and $\|P_2\|^{M,g}$ sub-combine into teams, for carrying out the actions. Some of such actions are possibly collective, because at least one of their participants is a team, not a singular individual.

Note that, in (11), P_1 and P_2, i.e., the variables denoting the witness sets, are no longer existentially quantified. Drawing from insights of (Schwarzschild 1996), the formula in (11) has to be interpreted both with respect to a model M and an assignment g, that provides a value for the free variables.

A g's assignment that makes (3) true in context (4) is the following:

(12) $\|P_1\|^{M,g} = \{c_1, c_2, c_3\}$ $\|P_2\|^{M,g} = \{p_1, p_2, p_3, p_4, p_5\}$
$\|C\|^{M,g} = \{\langle c_1 \oplus c_2 \oplus c_3,\ p_1 \oplus p_2\rangle, \langle c_2 \oplus c_3,\ p_3 \oplus p_4\rangle, \langle c_3,\ p_5\rangle\}$

This paper precisely focuses on this architectural choice. As it will be explained below, g must be designed in order to take into account all pragmatic factors needed to identify the witness sets the sentence is about.

4 Local and Global Maximality Conditions

Let us mark Sher's maximality conditions, like (9) in boldface, as 'Local maximality conditions'. Their name comes from the fact that they require the nonexistence of cartesian products including *the one* generated by $\|P_1\|^{M,g}$ and $\|P_2\|^{M,g}$. In other words, witness sets are firstly selected, then they are maximized, i.e., it is asserted that no superset of them satisfying the formula exists.

The kind of maximization is the same as the one proposed by Robaldo (2011). The latter simply requires maximal covers, that are a generalization of maximal cartesian products. The crucial difference between the proposals of Robaldo (2011) and Sher (1997) is that the former selects witness sets via an assignment g rather than via an existential quantification.

As noted by Schein (1993), pp. 284–293, Sher's maximality conditions yield incorrect truth values when the model includes two or more isomorphic cartesian products satisfying the predication. Schein considers the evaluation of the following sentences:

(13) a. Exactly two dots are connected with exactly two stars.
b. At least two dots are connected with at least two stars.

in the following model:

The sentence (13.a) is false in Fig. 1, while it is easy to see that the corresponding formula[2] in Sher's is true, given that it requires the existence of *at least one* relevant cartesian product.

Sher (1990) tries to fix her formulae by introducing an additional clause that requires the uniqueness of the cartesian product. A "unique maximum" requirement on witness sets is also advocated by Winter (2001, pp. 233).

[2] That formula is obtained from (9) by simply replacing the quantifier condition "$3!_y(\text{star}'(y), P_2(y))$" with "$2!_y(\text{star}'(y), P_2(y))$".

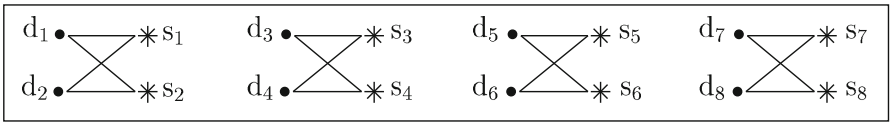


Fig. 1 A model for sentences (13.a–b)

However, Schein (1993) observes that such a uniqueness requirement leads to incorrect truth values when only M↑ quantifiers are involved. In particular, Schein provides evidence that both versions of Sher's are unable to state that sentences (13.a) is false in Fig. 1, while (13.b) is true.

In light of Schein's observations, (Landman 2000), and recently (Brasoveanu 2012), propose an alternative maximization, called here 'Global maximization'.

Landman (2000) represents[3] sentence (13.a) via the formula (14). Clauses in boldface are (Global) maximality conditions.

$$
\begin{aligned}
(14)\ \exists e \in^* \text{CONNECT} : \exists x \in^* \text{DOT} : |x| = 2 \wedge^* \text{Ag}(e) = x \wedge \\
\exists y \in^* \text{STAR} : |y| = 2 \wedge^* \text{Th}(e) = y \wedge \\
|^*\mathbf{Ag}(\bigcup\{e \in \mathbf{CONNECT} : \mathbf{Ag(e)} \in \mathbf{DOT} \wedge \mathbf{Th(e)} \in \mathbf{STAR}\})| = \mathbf{2} \wedge \\
|^*\mathbf{Th}(\bigcup\{e \in \mathbf{CONNECT} : \mathbf{Ag(e)} \in \mathbf{DOT} \wedge \mathbf{Th(e)} \in \mathbf{STAR}\})| = \mathbf{2}
\end{aligned}
$$

Formula (14) asserts the existence of a plural event *e* whose agent is a plural individual made up of two dots and whose theme is a plural individual made up of two stars.

The maximality conditions state that the *global* number of dots in the model must be exactly two (and so for the stars). In other words, contrary to what is done in Local maximization, the clauses in boldface do not refer to the same events and individuals quantified in the first row. Rather, they require that the number of dots connected to a star *in the whole model* is exactly two and so for the number of stars connected to a dot *in the whole model*. In other words, the witness sets are selected via an existential quantification, as in (Sher 1997), and the maximality conditions assert that these witness sets are *the* sets of all individuals participating in the main predicate's extension.

It is easy to see that, under these conditions, (13.a) comes out false in Fig. 1. Conversely, (13.b) comes out true, provided we substitute '=2' with '≥2' everywhere in the formula.

4.1 Are Local and Global Maximizations Effective Solutions?

We have seen that Sher's theory is problematic when the model includes multiple pairs of witness sets that satisfy the main predicate, e.g., Fig.1.

Landman (2000) and Brasoveanu (2012) propose then to use Global maximization, a solution that works, at least in the models analyzed by Schein. Nevertheless, it is rather easy to find examples where Global maximization does not predict the proper

[3] Landman (2000) does not consider Branching Quantifier readings. (14) is a cumulative reading among a set of exactly two dots and a set of exactly two stars. Cartesian products are only special instantiations of CONNECT's extension.

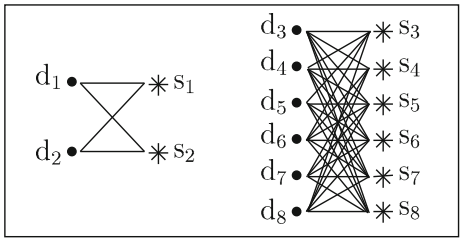


Fig. 2 A model for sentence (13.a)

truth values. It even seems that often such truth values are not so uncontroversially and mathematically determined. Consider the scenario in Fig. 2.

In our view, it cannot be felicitously said that (13.a) is false in Fig. 2. The visual effect obtained by the dense sub-structure on the right tend to "isolate" the one on the left, i.e., to see the latter separated from the rest. Such a visual effect does not occur in Fig. 1, where the sub-structures are completely isomorphic. On the other hand, the multiple availability of witness sets does not seem to confuse the reader for sentences involving M↑ quantifiers, perhaps because they are simpler to interpret (cf. Geurts and van der Silk 2005; Szymanik and Zajenkowski 2013).

In light of these considerations, relying on Global maximization appears to be a solution too rigid for being fully effective. The interpretation of multi-quantifier sentences does not always take into account all individuals in the model. Sometimes, the sentence's meaning is restricted to sub-groups of individuals, in case certain "pragmatic factors" occur either in the sentence or in the model.

However, Global maximization is intrinsically unable to handle such cases, *because it maximizes the main predicate's extension, not the selected witness sets*.

On the contrary, Local maximization is asserted *on* the selected witness sets, and so it is able to deal with both cases, i.e., both when the witness sets are *the* sets of all individuals participating in the main predicate's extension (Global reading), and when they are only subsets of the latter (Local reading).

Therefore, rather than modifying the way witness sets are maximized, a more flexible mechanism for *selecting* them should be provided. A solution could be the one proposed by Robaldo (2011). The 2-order variables occurring in the formulae are no longer existentially quantified. Rather, their value is provided by an assignment g. g is in charge of implementing the preference criteria, i.e., what we termed above as "pragmatic factors", involved in the witness sets' selection.

For testing how pragmatic criteria may be engaged in the interpretation, we conducted an online questionnaire in Italian. The statistical analysis of the results seems to confirm the hypothesis about the crucial role of pragmatics in quantifiers' interpretation, and *in what way* pragmatics is involved.

5 An Online Questionnaire on IS Readings

This section explains how the questionnaire was carried out, the set of trials that subjects are asked to evaluate, and our original predictions about them.

5.1 Participants

The data were collected via an online questionnaire that was made available by a http server. We used the social network Facebook to invite people to the questionnaire. The questionnaire displays pairs of Italian sentences and figures describing contexts with boys eating pizzas. Subjects had to decide whether the sentence was true given a selected figure. The questionnaire was filled out by 20,092 Italian participants (55.86 % female). Their mean age was 24 years (SD = 8).

In order to check whether the results of the questionnaire apply to languages other than Italian, we translated the questionnaire into three other languages (Polish, English, and German) for which we collected 809 additional answers. The additional results are presented in "Appendix". No significant differences were found in the answers provided when the questionnaire was presented in different languages, which suggests that our findings are language-independent. However, we know that such a generalized conclusion across languages is not viable, as there are known cases of quantifiers subject to different readings in different languages. In other words, further empirical investigations need to be carried out to explore the extent to which quantifiers differ in different languages. Further details on this issue may be found in "Appendix".

5.2 Materials and Procedure

The questionnaire included eight trials with target sentences and twelve fillers[4]. Participants were asked to evaluate one sentence at a time. Two target sentences were never adjacent in the sequence; one or more fillers were always displayed between them. The target sentences, listed below in (15), had the same syntactic structure of (3). They only varied with respect to the subject/object quantifiers:

(15) a. Esattamente tre ragazzi hanno mangiato esattamente tre pizze. (Trial 1)
(*Exactly three boys ate exactly three pizzas*.)
b. Esattamente un ragazzo ha mangiato esattamente una pizza. (Trial 2)
(*Exactly one boy ate exactly one pizza*.)
c. Meno di tre ragazzi hanno mangiato esattamente una pizza. (Trial 3)
(*Fewer than three boys ate exactly one pizza*.)
d. Più di tre ragazzi hanno mangiato la maggior parte delle pizze. (Trial 4)
(*More than three boys ate most pizzas*.)
e. Meno della metà dei ragazzi ha mangiato esattamente tre pizze. (Trial 5)
(*Fewer than half of the boys ate exactly three pizzas*.)
f. Esattamente due ragazzi hanno mangiato esattamente tre pizze. (Trial 6)
(*Exactly two boys ate exactly three pizzas*.)

[4] Fillers with obvious truth values (e.g., "In the figure, there are eight boys") were used to prevent subjects from using some simplified strategy that could only work with specific experimental target items. The full list of fillers is not reported in this paper, because their results were not stored in the database.

g. Più di cinque ragazzi hanno mangiato più di quattro pizze. (Trial 7)
(*More than five boys ate more than four pizzas*.)
h. Meno di tre ragazzi hanno mangiato esattamente una pizza. (Trial 8)
(*Fewer than three boys ate exactly one pizza*.)

Sentences have been chosen in the light of examples mentioned in the previous sections. We are interested to directly study claims from the literature on IS readings.

Three sentences involve non-M quantifiers only; they are isomorphic to the examples used by Schein for arguing contra Sher. Other three sentences involve a M↓ quantifier in the subject and a non-M quantifier in the object. They represent mixed cases that are not discussed by Schein, but that we consider anyway relevant for empirical research. Finally, two sentences involve M↑ quantifiers only.

Each sentence was associated with four figures, describing each a scenario with boys eating pizzas. All sentences but the two involving only M↑ quantifiers are evaluated in scenarios that include sub-structures of witness sets satisfying the main predicate. In other words, their evaluation in these scenarios is logically false under Global maximization, while it is true under Local maximization plus pragmatic identification of the witness sets. On the contrary, for the two sentences involving only M↑ quantifiers the opposite holds: they are true under Global maximization and false under Local maximization plus pragmatic identification of the witness sets.

Participants had to decide whether the sentence was true given a selected figure. The system selected the figure so that each of the four figures associated with a sentence was evaluated by an equal number of subjects (see below). Subjects provided an answer by pressing one of the corresponding buttons "Yes", "No", or "Don't know" ("Sì", "No", and "Non lo so", in Italian). Figure 3 shows a screenshot of the questionnaire. The Italian text "*Nella figura sottostante...*" translates into "*In the figure below...*". The sentence shown is the one of Trial 3 (*Fewer than three boys ate exactly one pizza*.).

At the beginning, subjects read a page of instructions that explained how to read the figures. Boys are connected to pizzas by means of lines that represent eating actions. If more boys are connected to a single pizza, the boys ate that pizza together, by cutting it into slices. Boys can have different colors. Boys with the same color are assumed to belong to the same football team.

Our predictions with respect to Fig. 3 were that subjects most likely would answer "Sì" ("Yes"). If they did it, it may be concluded that they identified a subgroup of fewer than two boys (presumably, the two ones in red) who ate a single pizza.

Of course, it may be argued that although the sentences in (15) are grammatically acceptable, they are very hard to find in everyday language. In other words, they may sound odd without a specific context. Nevertheless, as pointed out above, we are interested in directly studying claims from the literature on IS readings, and so we need to use sentences that are structurally similar to the ones used in Schein (1993), Sher (1997), Landman (2000), Robaldo (2011), and Brasoveanu (2012). That is why our instructions provide a natural context in which a sentence of the form "X boys ate Y pizzas" may be interpreted as "a group of X boys ate a group of Y pizzas". However, the possibility of judging a sentence as odd in a certain context is still accounted for, both in the questionnaire (subjects may press the "Don't know" button) and in the advocated logical theory (the assignment g may fail, meaning that it is unable to

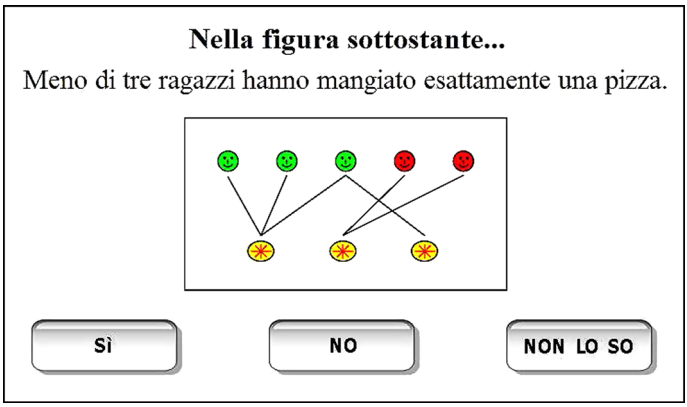


Fig. 3 A screenshot of the questionnaire

detect whether the model contains or does not contain suitable witness sets). It could also be argued that even sentences that appear to be linguistically valid could sound odd in certain contexts, and thus that this possibility should be always allowed.

5.3 Pragmatic Factors

The four figures associated with a sentence differ for two pragmatic factors. The role of the two pragmatic factors is to either favor or disfavor the desired (either Local or Global) interpretation. Figures may include additional minor pragmatic factors that may further induce the predicted reading. The two main pragmatic factors are:

(16) 1. *Color* All boys have the same color, or subgroups of boys are highlighted by different colors, as in Fig. 3 In our hypotheses, the latter tends to favor Local interpretation.
2. *Arrangement of the sub-structures*. Two alternative arrangements have been adopted in the figures: crossings and distance.
   (a) *Crossings* Connections (i.e., lines) between boys and pizzas do not cross between subgroups, or some connections in one group cross with some connections in another group. In our hypotheses, crossing disfavors Local interpretation.
   (b) *Distance* The distances between all boys are similar, as in Fig. 3, or subgroups of boys are spaced further apart. In our hypotheses, spacing favors Local interpretation.

The color/no-color alternation was tested within all target sentences. However, the crossing/no-crossing alternation was only tested within target sentences 2 and 3, whereas the distance/no-distance alternation was tested within target sentences 1, and 4–8.

Therefore, two separate analyses were performed. The analysis of the factors *color* and *distance* included target sentences 1, and 4–8, and the analysis of *color* and *crossing* included target sentences 2 and 3.

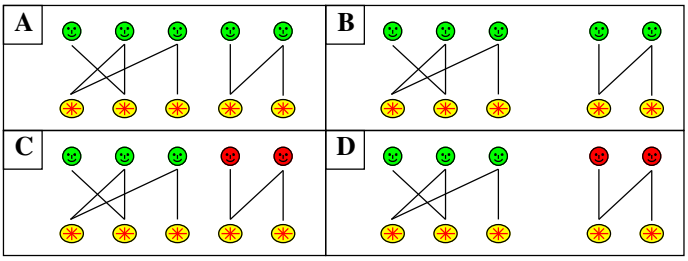
Given the two pragmatic factors, four scenarios are generated: one involving both of them, two ones involving either one of them, and one involving neither. The full list of scenarios is shown in the next subsection.

Each subject evaluates only one of the four figures associated with every trial. In order to guarantee that all 8×4 scenarios are evaluated by an equal number of subjects, we built tuples of twenty trials, each corresponding to a different combination. Nevertheless, we did not generated all 4^8 combinations. We wanted to avoid, for instance, the tuple where *all* its trials do not involve any pragmatic factor. Thus, we imposed some constraints on tuples' generation. They all include a sequence of trials where colored boys are alternated. For instance, in half of the tuples Test1 involves colored boys, Test2 involves non-colored ones, Test3 involves colored ones, etc. The other half also features that alternation, but it starts by showing non-colored boys: Test1 involves non-colored boys, Test2 involves colored ones, Test3 involves non-colored ones, etc. Then, all possible combinations[5] of (non-)arrangement of the substructures are generated. Therefore, we obtained $2 \times 2^8 = 512$ tuples. The order of the tuples was then randomized and stored in the database. Once a subject connected to the server, a tuple of trials among the ones that have been evaluated by fewer persons was selected. Thus, each of the 512 tuples was evaluated by an (almost) equal number of subjects.

5.3.1 The Full List of Figures

In this subsection, we report the four scenarios associated with each of the eight target sentences. Furthermore, we explain the additional pragmatic factors inserted therein and our predictions of the results.

Trail 1 Esattamente tre ragazzi hanno mangiato esattamente tre pizze (*Exactly three boys ate exactly three pizzas*)

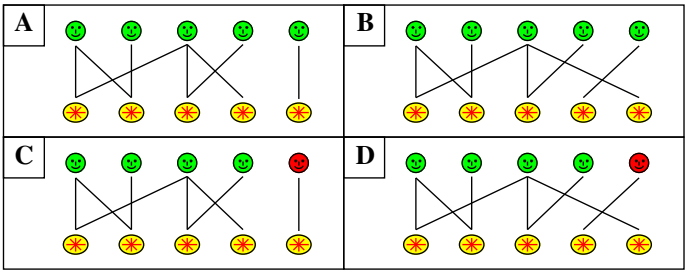


As pointed out above, three sentences of the questionnaire involve non-M quantifiers only. They are three variants of Schein's example illustrated in Fig. 1, that is true under Global maximization but false under Local one. The target sentence of Trial 1 is one of them. Trial 1 includes the same Exactly-n quantifier both in the subject and in the

[5] We privileged the pragmatic factor about color over the other one. In our view, the former mostly favor Local reading. Thus, it is more important to trial its effect. In the light of this, we think it is fine to leave tuples that include only trials without relevant arrangements of the items.

object, as sentence (13.a). However, the four scenarios differ from Fig. 1 because there is a single sub-structure satisfying the main predication, rather than multiple ones. We predicted that most subjects would answer "Yes" in the present trial.

Trail 2 Esattamente un ragazzo ha mangiato esattamente una pizza (*Exactly one boy ate exactly one pizza*)

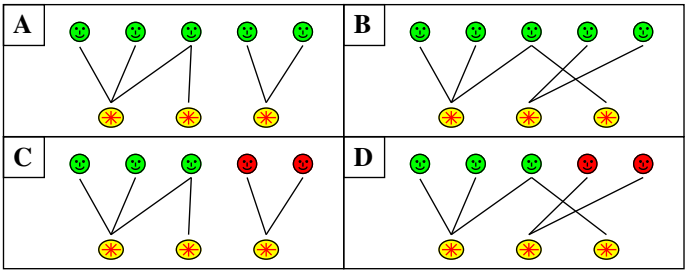


Also the sentence used in Trial 2 involves non-M quantifiers only. Trial 2 shares with Trial 1 the non-occurrence in the model of multiple sub-structures satisfying the main predication and the use of the same quantifier both in the subject and in the object. However, we predicted a number of "Yes" answers greater than those of Trial 1 because the Exactly-n quantifier is "Exactly one". The latter seems to have a strong pragmatic preference towards Local readings. Therefore, in our view subjects most likely would focus on the single boy connected to a single pizza, that occurs in the figures.

Even if the crossing in figures (B) and (D) should disfavor the identification of this sub-structure, we predicted that "Exactly one" is stronger than the crossing pragmatic factor, i.e., to collect an high number of "Yes" answers also for (B) and (D).

Trial 2 is considered "simpler" than Trial 1. Therefore, it has been inserted in second position. We did not want to show the one that we consider the "simplest" trial as first. Trial 1 is considered of "medium complexity". The trial involving non-M quantifier only that we consider as the most complex is shown in sixth position (see below).

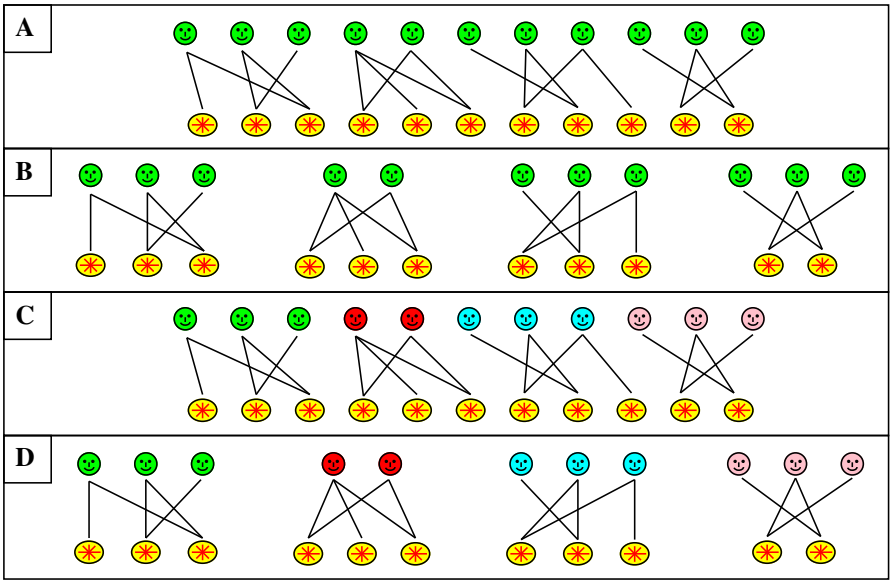
Trail 3 Meno di tre ragazzi hanno mangiato esattamente una pizza (*Fewer than three boys ate exactly one pizza*)



Trial 3 is similar to Trial 2. The only difference is that it involves a M↓ quantifier in the subject, in place of "Exactly one". As in Trial 2, we predict that 'Exactly one" should strongly favor the identification of the sub-structure of two boys and a pizza,

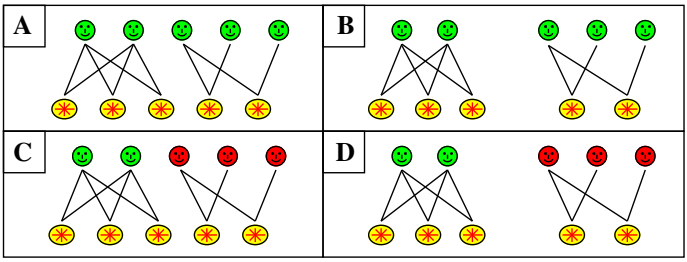
even in scenarios (B) and (D), where the lines starting from the two boys cross with another line.

Trail 4 Più di tre ragazzi hanno mangiato la maggior parte delle pizze (*More than three boys ate most pizzas*)



The sentence used in Trial 4 involves two M↑ quantifiers. In Schein's theory, i.e., under Global maximization, the sentence is true in scenarios (A)–(D). On the other hand, we predict that most subjects would answer "No", except perhaps in scenario (A). In the latter, we predict that they would most likely consider the eleven boys as a single group. If such results are met, it may be concluded that neither the truth values of M↑ quantifiers are so deterministic, despite their alleged simplicity from a cognitive point of view.

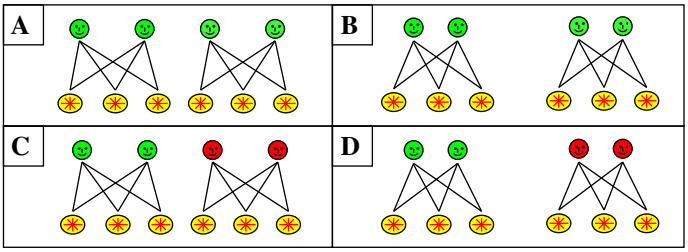
Trail 5 Meno della metà dei ragazzi ha mangiato esattamente tre pizze (*Fewer than half of the boys ate exactly three pizzas*)



Trial 5 is a variation of Trial 3. Both trials involve a M↓ quantifier in the subject and a non-M quantifier in the object. In our view, their level of cognitive complexity is

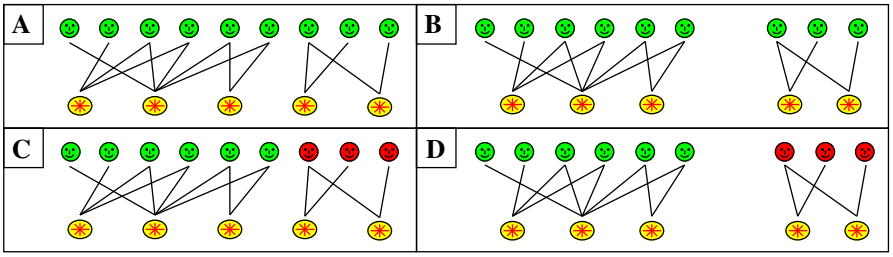
almost the same. In Trial 5, the quantifier in the subject is a proportional quantifier, which should be simpler to interpret than the counting quantifier used in the subject of Trial 3. On the other hand, the sentence of Trial 3 involves the quantifier "Exactly one" in the object. In our assumptions, "Exactly one" features a pragmatic preference towards the identification of sub-structures stronger than the one of "Exactly three".

Trail 6 Esattamente due ragazzi hanno mangiato esattamente tre pizze (*Exactly two boys ate exactly three pizzas*)



Trial 6 is the third and last test that uses non-M quantifiers only. It represents a third variant of the example used by Schein for arguing contra Sher. The only difference with respect to the latter is that the Exactly-n quantifier in the subject is different from the Exactly-n quantifier in the object. On the other hand, both the four scenarios in Trial 6 and the model in Fig. 1 include multiple isomorphic sub-structures, satisfying the main predication, where every boy is connected to every pizza. For this reason, as pointed out above, we predict that the results for Trial 6 would have a lower percentage of "Yes" answers than Trial 1 and Trial 2.

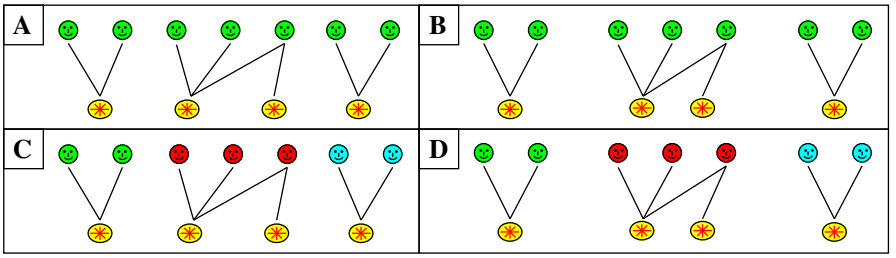
Trail 7 Più di cinque ragazzi hanno mangiato più di quattro pizze (*More than five boys ate more than four pizzas*)



Trial 7 is the second trial of the questionnaire that involves two M↑ quantifiers, the other one being Trial 4. With respect to Trial 4, Trial 7 involves fewer boys and fewer pizzas and its scenarios contain only two sub-structures of boys and pizzas. Moreover, both quantifiers in Trial 7's sentence are numerical quantifiers. Their interpretation should be more deterministic than the one of "Most", used in Trial 4's sentence.

For these reasons, we predicted that subjects would globally interpret Trial 7's sentence most likely than in Trial 4. In other words, we expected a greater percentage of "Yes" answers in the results of Trial 7 than in those of Trial 4.

Trail 8 Meno di tre ragazzi hanno mangiato esattamente una pizza (*Fewer than three boys ate exactly one pizza*)



The final trial, like Trial 3 and Trial 5, involves a M↓ quantifier in the subject and an Exactly-n quantifier in the object. The Exactly-n quantifier in the object is "Exactly one", that should favor the identification of sub-structures. On the contrary, each scenario in Trial 8 includes two isomorphic sub-structures satisfying the main predication, a feature that should disfavor Local maximization. The target sentence is the same used in Trial 3. Nevertheless, we predict to register a greater number of "No" answers in Trial 8's scenarios than in Trial 3's ones, because the former involve more boys and pizzas than the latter and includes two isomorphic sub-structures satisfying the main predication.

6 Results

The following analyses concern the effects of *color* (no-color vs. color), *crossing* (no-crossing vs. crossing), and *distance* (no-distance vs. distance) on the probability of a local interpretation. The data were analyzed by means of logistic regression, because the outcome variable is binary (local interpretation vs. global interpretation). More specifically, we fitted logistic *linear mixed-effects* models, to account for random variation due to *individual differences* and *test items*. Traditional analysis of variance (ANOVAs) cannot account for multiple sources of random variation.

We fitted separate models for the factors *distance* and *crossing*, as *distance* was manipulated for sentences 1, 4, 5, 6, 7, and 8, and *crossing* for sentences 2 and 3. Both models include the factor *color*, as color was manipulated across all sentences.

Figure 4 depicts the mean proportion of local interpretations across participants. It clearly shows an effect of color. In fact, the main effect of color is significant; for target sentences 1, 4, 5, 6, 7, and 8 (Fig. 4a), $\chi^2(1) = 229.0$, $p < .001$, and for target sentences 2 and 3 (Fig. 4b), $\chi^2 = 29.0$, $p < .001$. On average, the proportion of local interpretations was highest if the groups of boys were depicted in distinct colors. The proportion of local interpretations was slightly lower if there were no distinct colors indicating separate groups of boys.

On visual inspection, there does not seem to be an effect of distance (Fig. 4a). However, the main effect of distance is significant, $\chi^2 = 5.8$, $p = .016$, probably due to our large sample size. The proportion of local interpretations was just slightly higher if the groups of boys were depicted with greater distance between them. The mean difference in the proportion of local interpretations, between the no-distance versus

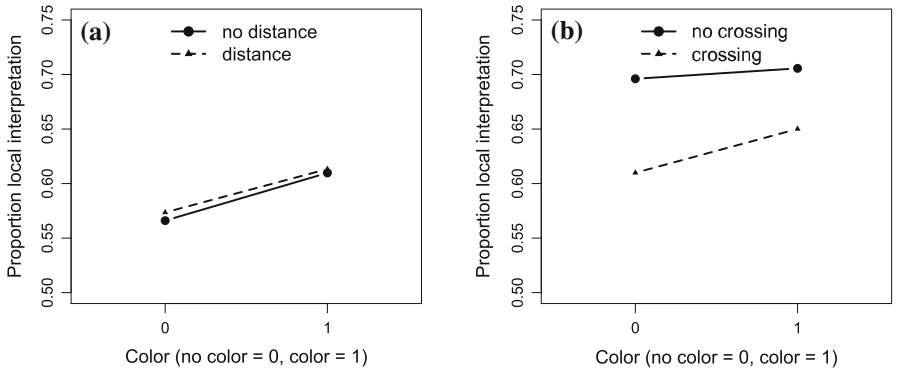


Fig. 4 Mean proportion of local interpretations, and the effects of distance (**a**), crossing (**b**), and color (**a** and **b**). Standard errors were too small to depict

distance conditions, was 0.6 %. As can be seen in Fig. 4, there was no significant interaction between the factors color and distance.

Figure 4b clearly shows an effect of crossing. The mean proportion of local interpretations decreases if the connections between boys and pizzas cross between subgroups of boys. This main effect is significant, $\chi^2 = 224.6$, $p < .001$.

There is also a significant interaction between color and crossing, $\chi^2 = 8.9$, $p = .003$. The positive effect of color on Local Readings was strongest if the scenarios (i.e., pictures of boys eating pizzas) included crossings between subgroups of boys. In other words, the presence of crossings between subgroups of boys disfavors a Local Reading, but that effect diminishes if the subgroups of boys are highlighted by distinct colors.

6.1 Sentence-by-Sentence Analyses

This subsection shows and discusses how well the predictions fit each individual test trial. For each trial, we present a table that reports the number of local (i.e., "Yes" answers) and global (i.e., "No" answers) interpretations. We perform Chi-squared test to compare within-sentence effects of color, distance, and crossing, and to compare the number of local / global interpretations between sentences. All p values reported in this section were adjusted by means of the Bonferonni-Holm correction to account for the family-wise error-rate.

Trail 1 Exactly three boys ate exactly three pizzas The results of Trial 1 do not meet our predictions about it. Most subjects interpreted the sentence globally, i.e., by considering all boys as a whole group. However, the proportion of local interpretations is significantly greater for scenarios (C) and (D) than for scenarios (A) and (B), which involve the pragmatic factor Color, $\chi^2(1, N = 19{,}054) = 92.44$, $p < .001$.

Scenarios (B) and (D) involve the pragmatic factor Distance, in contrast to scenarios (A) and (C). However, the number of local interpretations does not significantly differ between the former and latter scenarios (Table 1).

Table 1 Evaluation of Trial 1 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	1,522	3,182	254	32.35
B	1,477	3,289	260	30.99
C	1,789	3,016	257	37.23
D	1,884	2,895	267	39.42

Table 2 Evaluation of Trial 2 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	3,487	1,432	152	70.89
B	2,980	1,886	171	61.24
C	3,337	1,467	162	69.46
D	3,153	1,693	172	65.06

Trial 1 is the only trial where our hypotheses are not confirmed. The other two trials involving an Exactly-n quantifier both in the subject and in the object (Trial 2 and Trial 6) do have a greater number of "Yes" answers than "No" answers.

Trial 1 is the first trial shown to the subjects, after the instruction page (and three fillers). Several explanations appear to be available for justifying its results. It could be that Trial 1 features an intrinsic preference towards Global maximization. Or, it could be that subjects are somehow confused at the beginning of the questionnaire, and that confusion induced them to press "No" by following an instinctive default preference towards Global maximization.

Trail 2 Exactly one boy ate exactly one pizza The results partially meet our predictions for Trial 2. Most subjects identified the single boy connected with a single pizza in each scenario, and considered the sentence true in each scenario. Note that scenarios (A) and (C) have a significantly greater proportion of local interpretations (i.e., "Yes" answers) than scenarios (B) and (D), $\chi^2(1, N = 19{,}435) = 107.91$, $p < .001$. The latter are the two scenarios involving the pragmatic factor Crossing that, as we expected, induces participants to consider the boys as a whole group.

The number of local interpretations did not significantly differ between scenarios that involved the pragmatic factor Color ((C) and (D)) and those that did not ((A) and (B)) (Table 2).

Trail 3 Fewer than three boys ate exactly one pizza The results of Trial 3 fully meet our expectations. The scenarios that include the pragmatic factor Crossing ((A) and (C)) have a significantly lower number of local interpretations than the scenarios without Crossing, $\chi^2(1, N = 18{,}884) = 107.6$, $p < .001$ (Table 3).

Scenarios that include the pragmatic factor Color ((C) and (D)) have a significantly greater number of local interpretations than the scenarios without Color, $\chi^2(1, N = 18{,}884) = 31.24$, $p < .001$.

Table 3 Evaluation of Trial 3 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	3,187	1,482	301	68.26
B	2,852	1,844	318	60.73
C	3,444	1,362	263	71.66
D	3,061	1,652	326	64.95

Table 4 Evaluation of Trial 4 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	1,680	2,176	1,229	43.56
B	1,729	2,023	1,271	46.08
C	1,528	2,198	1,258	41.00
D	1,754	2,023	1,223	46.43

Trail 4 More than three boys ate most pizzas For Trial 4, we registered a percentage of "No" greater than the one of "Yes" in all four scenarios. We think this is a great result in favor of Local maximization. Although the sentence used in Trial 4 involves two M↑ quantifiers, subjects do not tend to consider the boys in the figures as a whole group (Table 4).

Another interesting result is the relevant number of "Don't know" answers, much greater than the one registered in the other trials. In our view, several subjects felt unable to evaluate the sentence because they found the scenarios too complex, perhaps because of the high number of items and lines occurring therein.

By using an assignment g for identifying the witness sets, as proposed by Robaldo (2011), it is possible to emulate Trial 4 subjects' behavior. The function g can fail, because it does not have enough knowledge for identifying the witness sets in the context. Of course, such an outcome corresponds to a "Don't know" answer. A dialogue system could then overcome the problem by asking the interlocutor to provide further knowledge for properly identifying the sets of entities he/she refers to.

The pragmatic factor Color did not significantly increase the number of local interpretations, but the factor distance did have a positive effect on the number of local interpretations, $\chi^2(1, N = 15111) = 23.73$, $p < .001$.

Trail 5 Fewer than half of the boys ate exactly three pizzas The results of Trial 5 are surprisingly good with respect to our predictions. We did not expect the proportion of "Yes" answers to be greater than the proportion of "Yes" answers acknowledged for Trial 2, which involves two occurrences of the "Exactly one" quantifier. However, the proportion of local interpretations is significantly higher for Trial 5 than for Trial 2, $\chi^2(1, N = 37{,}719) = 132.07$, $p < .001$. Thus, it seems that the truth values of multi-quantifier sentences with mixed monotonicity do not easily reconcile with Schein's assumptions (Table 5).

However, the pragmatic factors Color and Distance are not significant.

Table 5 Evaluation of Trial 5 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	3,221	1,304	463	71.18
B	3,283	1,276	437	72.01
C	3,330	1,296	462	71.98
D	3,355	1,219	446	73.35

Table 6 Evaluation of Trial 6 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	2,713	2,069	281	56.73
B	3,182	1,611	252	66.39
C	3,217	1,499	250	68.21
D	3,388	1,350	280	71.51

Table 7 Evaluation of Trial 7 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	2,063	2,357	542	46.67
B	2,125	2,372	525	47.25
C	1,895	2,701	506	41.23
D	1,800	2,708	498	39.92

Trail 6 Exactly two boys ate exactly three pizzas Similarly to Trial 5, we did not expect a so high number of "Yes" answers for Trial 6. The latter is rather close to Schein's example in Fig. 1, the only relevant difference being the use of two different Exactly-n quantifiers in the subject and in the object. For this reason, as said above we predicted that subjects would tend to interpret Trial 6 under Global maximilization more likely than in Trial 1. The results attested the opposite. However, this effect might be biased a bit, because of the first-sentence effect (Table 6).

As hypothesized, the pragmatic factors Color and Distance both have a significant positive effect on the proportion of local interpretations, $\chi^2(1, N = 19{,}029) = 144.96$, $p < .001$ and $\chi^2(1, N = 19{,}029) = 88.86$, $p < .001$, respectively.

Trail 7 More than five boys ate more than four pizzas The target sentence used in Trial 7 involves two M↑ quantifiers, like the one used in Trial 4. We got results similar to the ones of the latter. They seem to provide further evidence about the not so easy evaluation of sentences involving M↑ quantifiers only. Note that the numbers of "Don't know" answers are lower than the ones registered for Test4. Therefore, Trial 7's scenarios appear to be less confusing than Trial 4's ones, probably because they involve less boys, pizzas, and lines (Table 7).

Table 8 Evaluation of Trial 8 in scenarios (A)–(D)

Scenario	Yes	No	Don't know	Yes (%)
A	3,237	1,433	392	69.31
B	3,173	1,473	400	68.30
C	3,262	1,325	384	71.11
D	3,235	1,395	383	69.87

Quite surprisingly, the pragmatic factor Color had a significant negative effect on the proportion of local interpretations, opposite to the positive effect of Color for the other sentences, $\chi^2(1, N = 18{,}021) = 74.25$, $p < .001$. The pragmatic factor Distance did not significantly increase the proportion of local interpretations.

Trail 8 Fewer than three boys ate exactly one pizza The results of Trial 8 partially meet our predictions. Note that the proportion of "Yes" answers is almost the same in each scenario. The pragmatic factor Color did affect the evaluation of the sentence, but the factor Distance did not. The proportion of local interpretations is significantly higher for scenarios that included Color ((C) and (D)) than for scenarios that did not ((A) and (B)), $\chi^2(1, N = 18{,}533) = 6.13$, $p = .013$.

Nor the occurrence in the models of two isomorphic witness sets satisfying the main predication seems to induce subjects' interpretation towards Global maximization (Table 8).

This sentence has a significantly greater proportion of local interpretations that the sentence in Trial 3, $\chi^2(1, N = 37{,}417) = 44.34$, $p < .001$.

7 Discussion and Outlook

Although our expectations has been, by and large, met, it is clear that eight trials are not enough for saying the last word on the role of pragmatics in the interpretation of multi-quantifier sentences and/or the proper formalization of such sentences. We designed the scenarios to support the existence of Local readings, i.e., to provide counter-examples to Schein's claims. And, indeed, it turned out that Local readings are available, under certain circumstances. Therefore, the logical formulae must incorporate a treatment of pragmatic factors that enables them. In Sect. 6.1, some alternatives are considered.

On the other hand, although the trademark feature used to distinguish the trials has been monotonicity (three trials involving non-M quantifiers only, three trials involving a M↓ quantifier in the subject and a non-M quantifier in the object, two trials involving M↑ quantifiers only), from questionnaire's results it seems that new relevant features are involved in quantifiers' interpretation. Section 7.2 outlines the main ones. They will be addressed in our future questionnaire/experiments.

7.1 Formalizing IS Readings

This subsection outlines some of the possible solutions for incorporating an account of pragmatic factors within the logical formulae.

A good survey of the alternative solutions is given by Stanley and Szabò (2000). There are basically two approaches to the problem, termed by Neale (1990) and Reimer (1998) as the *explicit* and the *implicit* approach. The former deals with quantifier domain restrictions as if they were ellipses. In the latter, formulae are evaluated into local sub-models, considered more salient than the whole model with respect to a certain interpretation. Stanley and Szabò (2000) argue that both approaches seem to be inadequate.

In particular, "Model-theoretic" solutions belonging to the explicit approach require the definition of model theory rules relating the truth conditions of the two (nested) models, e.g., some rules like the following:

(17) For any formula Φ and any pair of models M' and M, such that $M' \subseteq M$ and Φ is more salient in M' than in M it holds:

$$\Phi^{M'} \models \Phi^{M}$$

With respect to the scenarios of our questionnaire, (17) allows to states that if a certain sentence is true for a sub-cover of boys and pizzas, then it is true in the whole figure.

Stanley and Szabò (2000) argue that (17) works for single-quantifier sentences, but it is not fine-grained enough for properly handling sentences including more than one quantifier. They provide the following counter-example to (17):

(18) Every sailor waved to every sailor.

Sentence (18) may be true in a context where every sailor on the ship waved to every sailor on the shore. (17) is unable to capture these truth conditions, because the predicate $sailor$ denotes a single and fixed set $\|sailor\|^{M'}$ in the local sub-world.

Furthermore, as discussed by Robaldo (2010b), handling maximization in terms of the model theory rules complicates reasoning. The inference theory should take into account constraints on such rules, a solution that seems at odds with standard literature on automatic reasoning. In addition, (17) would require the introduction of additional rules for avoiding the assertion in M of inferred clauses that must only hold in M'.

In light of these observations, both Stanley and Szabò (2000) and Robaldo (2011) advocate the use of variables able to denote different sets of entities. Thus, the two occurrences of 'Every sailor' in (18) could refer to different sets of sailors, and, of course, what is inferred for either set does not necessarily hold also for the other one.

As illustrated above, in Robaldo (2011) the interpretation of quantifiers is fully devolved upon an assignment g, in line with (Schwarzschild 1996). g provides a value for every 2-order variable occurring in the formulae. g's selection may fail, and so the sentence is either taken to be false (by default) or at least "odd" in that context (cf. comments to Trial 4 in Sect. 5.1 above). Independently of the fact that the witness sets are identified either locally or globally, explicit maximality conditions apply to them, in order to trigger appropriate inferences via standard techniques, e.g., resolution, as shown by Robaldo (2010b). Such inferences assert new clauses that hold only for the sets identified by g, e.g., for the set of sailors on the ship, but not necessarily for other (super-)sets, e.g., the set of sailors on the shore or the whole set of sailors in the universe.

A similar solution has been proposed by Brasoveanu (2012). The author acknowledges the existence of both the Global and the Local reading for the sentence "Exactly

three boys saw exactly five movies". Nevertheless, he states: "it is not clear that this (the Local reading) is even a possible reading for the sentence" and so he develops a logical account handling Global readings only. Brasoveanu's account is grounded on a maximization operator σ that takes in input a conjunction of predicates and returns the (globally) maximal sums satisfying the cumulative reading of that conjunction.

After the maximal sums have been so identified, they are required to satisfy the cardinality constraints. For instance, with respect to Brasoveanu's sample sentence, they are required to include exactly three and exactly five individuals.

In order to handle Local readings in Brasoveanu (2012)'s, it seems sufficient to re-define the operator σ, making it able to select *local* sums. On the other hand, the crucial difference between Robaldo's and Brasoveanu's is that in the latter the cardinality requirements are evaluated *after* the operator σ has identified the witness sets, while in the former they contribute to g's selection. In other words, assuming, by default, a collaborative speaker, g looks in the context for local/global sums of individuals that satisfy *every* clause representing the meaning of speaker's utterance, both predicates and the cardinality constraints conveyed by quantifiers.

7.2 Planning New Experiments

Let us briefly comment on our empirical findings and acknowledge the need for proper psycholinguistic experimentation. The main goal of our empirical research was to provide evidence for local interpretation of multi-quantifier sentences. Moreover, we aimed to show that pragmatic aspects of the situation may influence that interpretation. And indeed, our research has shown that people sometimes interpret multi-quantifier sentences locally and that this tendency depends on some pragmatic factors. Still our questionnaire does not allow us to draw any rigorous conclusions on humans' interpretation process of multi-quantifier sentences since in experiments with such complex natural language material there might be many factors influencing subjects' judgements. We have seen that our data are congruent, for example, with Robaldo's proposal, but they are not specific enough to imply a concrete comprehension theory. The results just prove that every good theory of quantification should be able to take into account pragmatic factors influencing interpretation.

In the proper psycholinguistic experiment we would need to control many factors. For instance, consider our Trial 4, where people were supposed to evaluate the following sentence: "More than three boys ate most pizzas". In fact, in all our pictures all the pizzas were eaten. Therefore, according to the Gricean rules, a collaborative speaker would rather say: "More than three boys ate all pizzas." That observation can potentially explain the high number of "Don't know" answers in Trial 4. In the full fledged experiment such factors should be taken into account as well.

As our goal was much more modest than designing a proper theory of quantifier comprehension, namely providing evidence in favor of pragma-semantic approach to quantification, we did not have to control such subtle factors. Even if many subjects found the sentences of Trial 4 infelicitous and answer "Don't know", still there was a significant fraction of participants that interpreted the sentences locally.

Moreover, several cognitive experimental results showed that many factors may affect the quantifier interpretation. In our questionnaire, we practically controlled only monotonicity, but the future experiment should also take into account other variables that turned out relevant in many studies, for instance, computational complexity of quantifiers (Szymanik 2009; Szymanik and Zajenkowski 2009) or their fuzziness (Sanford and Paterson 1994).

One way to rigorously design simpler experiments investigating the influence of pragmatic factors on quantifier interpretation would be to focus on single-quantifier sentences. We expect that in that case one could clearly observe the pragmatic influence, and the preference of local interpretations in the speakers' judgements. We admit that it would be a natural order to start with simple sentences before continuing to multi-quantifier constructions, like those studied in the paper or others, e.g., reciprocal sentences of the form "Exactly six boys shared pizza with each other". However, we decided to focus on Independent Set readings as our goal was to directly contribute to the literature on scopeless readings. We believe that we achieved our aim by pointing out necessary improvements in the theory of complex quantifier sentences. Additionally, we raised general questions about the semantic-pragmatic interface in quantifier comprehension that need to be tackled in future research.

7.3 Implementation

The pragmatic factors affect the interpretation of the sentences. Therefore, the formulae representing meaning should take into account pragmatic preferences. Thus, the formulae would achieve the flexibility needed to denote the proper truth values in different contexts and under different (subjective) perspectives.

Possible solutions appear to be proposals by Robaldo (2011) or a slight revision of the framework suggested by Brasoveanu (2012). The former, drawing from the work of (Schwarzschild 1996), proposes to implement all pragmatic factors within an assignment g. The latter provides a value, i.e., a witness set, for every 2-order variable occurring in the formula. On the other hand, the identification may fail, and so the sentence is either taken to be false (by default) or at least "odd" in that context.

Independently of the fact that the witness sets are identified either locally or globally, explicit maximality conditions apply to them, in order to trigger appropriate inferences (cf. Robaldo 2010b). In other words, clauses of (Robaldo 2011)'s formulae must not be seen as "constraints that need to be satisfied in order to detect if a certain sentence is true in a certain context", but rather as "asserted facts about the identified witness sets, that could be used to infer new knowledge about them".

Our future works will be devoted to the implementation of the function g. In light of above discussion, g's design should not follow absolute criteria. A more promising solution would be to base g's outcome on a statistical model, that can be learned and updated. Different instantiations of that model would correspond to different preference criteria.

In order to devise such a statistical model, we plan to conduct further experiments to analyze how other pragmatic factors, like monotonicity, affect the interpretation. As we already noted in the result section the effect of distance is very small and we

realize that its significance may be mostly due to the large sample size. We think it can become more important for particular monotonicities. This is an extra reason to further explore the role of monotonicity, and its interaction with other pragmatic factors.

Another interesting open problem is the question of the computational complexity of IS readings. Cumulative and collective readings, being special cases of IS readings, have been recently studied from that perspective. Szymanik (2010) has shown that Cumulative readings are tractable (PTIME computable quantifiers are closed on Cumulative reading). Kontinen and Szymanik (2008) and Kontinen and Szymanik (2011) have proved that, on the contrary, Collective readings can be highly intractable (e.g., the collective reading of proportional quantifiers is not definable in the second-order logic). By studying the computational complexity of IS readings those results could be generalized. Especially, it would be interesting to compare computational complexity of IS readings under Local and Global maximization principles. Such analysis could provide additional arguments in favor of cognitive and linguistic plausibility of one of the interpretations (cf. Szymanik 2009, 2010; Mostowski and Szymanik 2012).

8 Conclusions

In this paper, we reported the results of an online questionnaire designed to study the subjects' interpretations of multi-quantifier sentences for which several approaches in the literature identified precise and mathematically-determined truth values. One of those is the approach originally advocated by Schein (1993), termed here as "Global interpretation", where the truth values are deterministically computed from the cardinalities of the sets of *all* entities occurring in the context and the monotonicities of the quantifiers.

Our experiments were built with the intention of falsifying the truth values predicted by the Global interpretation. The scenarios were enriched with some pragmatic factors. We predicted that they induce subjects to consider sub-teams of boys and pizzas in the scenarios, rather than *all* boys and pizzas.

Even though, given a huge ambiguity of the presented sentences, the results of the questionnaire does not support unequivocally any concrete formal semantic representation, they show that the assumptions lying behind Global interpretation are not necessarily empirically adequate. Therefore, note that we are not claiming that Schein's observations are wrong. There is indeed a general tendency to consider all individuals in the model. However, in our view, such a general tendency should be implemented as a kind of default rule that triggers only if it is not overridden by stronger pragmatic factors.

We hope that our work not only throw some light on the semantic of natural language sentences, contributing empirical observations to the the theoretical discussion, but also illustrates how the experimental methods may be fruitfully combined with logical investigations in order to discover the underlying structures of natural language.

Acknowledgments The authors would like to thank Lucas Champollion for suggestions to the previous versions of the online questionnaire and an anonymous reviewer for fruitful comments. BM and JS were supported by a Vici Grant NWO-277-80-001. JS also acknowledges NWO Veni Grant 639.021.232.

Results of the Questionnaire in Polish, English, and German

As mentioned at the beginning of Sect. 4, in order to check whether the results of the questionnaire apply to languages other than Italian, we translated the questionnaire into three other languages (Polish, English, and German). 809 additional participants answered to the questionnaire: 415 Polish, 305 English, and 89 German native speakers.

Table 9 reports the answers of the 809 non-Italian subjects.

Table 9 Evaluation of the eight trials by Polish, English, and German native speakers

Trial-scenario	Yes	No	Don't know	Yes (%)
Trial 1-A	72	139	10	34.12
Trial 1-B	63	126	9	33.33
Trial 1-C	79	127	7	38.34
Trial 1-D	70	93	14	42.94
Trial 2-A	137	61	7	69.19
Trial 2-B	110	68	7	61.79
Trial 2-C	144	63	4	69.57
Trial 2-D	140	62	6	69.31
Trial 3-A	143	154	14	72.59
Trial 3-B	119	78	11	60.41
Trial 3-C	136	52	9	72.34
Trial 3-D	128	51	14	71.51
Trial 4-A	66	71	49	48.17
Trial 4-B	71	78	55	47.65
Trial 4-C	62	86	55	41.89
Trial 4-D	71	92	53	43.55
Trial 5-A	129	41	23	75.88
Trial 5-B	143	62	21	69.75
Trial 5-C	116	51	20	69.46
Trial 5-D	138	48	17	74.19
Trial 6-A	117	74	11	61.26
Trial 6-B	134	42	12	76.14
Trial 6-C	135	70	10	65.85
Trial 6-D	137	56	11	70.98
Trial 7-A	82	100	25	45.05
Trial 7-B	94	89	29	51.37
Trial 7-C	58	84	22	40.84
Trial 7-D	80	121	25	39.80
Trial 8-A	126	65	14	65.97
Trial 8-B	127	45	13	73.84
Trial 8-C	128	59	19	68.45
Trial 8-D	138	55	20	71.50

Given the results shown in Table 9, it seems that language does not affect the interpretation of the sentences in our questionnaire. The proportions are by and large the same found for Italian. Nevertheless, we cannot ignore the fact that the meaning of some quantifiers differ across languages and that we cannot expect to obtain the same meaning by translating quantifiers literally.

For instance, with respect to the sentence in Trial 4 ("Più di tre ragazzi hanno mangiato la maggior parte delle pizze", translated in the English version of the questionnaire as "More than three boys ate most pizzas"), an anonymous reviewer suggested that in English "most" may have a proportional or a relative reading.

According to our informants and the available literature, e.g. (Hackl 2009; Krasikova 2011; Szabolcsi 2013), the proportional-relative ambiguity does not exist in English. The literature assumes that "most books" by itself is always proportional and "the most books" is always relative.

Indeed, such ambiguity does occur in German, but not in Italian or Polish. In our view, such potential ambiguity could not significantly affect our overall results. It seems most likely that English subjects interpreted "most" as a proportional determiner, especially, as it was explained very clearly in the instructions that trial sentences have the structure "X boys ate Y pizzas" and must be interpreted as "a group of X boys ate a group of Y pizzas". Moreover, this is the fourth trial, which took place before subjects had evaluated three syntactically similar trials with clearly proportional interpretation.

To conclude, although our findings appear to be language-independent, in this paper we should consider reliable only the results for the Italian version of the questionnaire, while the evaluation of the trials in other language would need further empirical analyses.

References

Alshawi, H. (1992). *The core language engine*. Cambridge, MA: MIT Press.

Beck, S., & Sauerland, U. (2000). Cumulation is needed: A reply to winter (2000). *Natural Language Semantics*, *4*(8), 349–371.

Brasoveanu, A. (2012). Modified numerals as post-suppositions. *Journal of Semantics*, *30*(1).

Cooper, R. (1983). *Quantification and syntactic theory*. Dordrecht: D. Reidel.

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, *21*, 159–210.

Diesing, M. (1992). *Indefinites*. Cambridge, MA: MIT Press.

Geurts, B., & van der Silk, F. (2005). Monotonicity and processing load. *The Journal of Semantics*, *22*(17).

Gierasimczuk, N., & Szymanik, J. (2009). Branching quantification versus two-way quantification. *The Journal of Semantics*, *4*(26), 329–366.

Hackl, M. (2009). On the grammar and processing of proportional quantifiers: Most versus more than half. *Natural Language Semantics*, *17*(1), 63–98.

Heim, I. (1982). *The semantics of definite and indefinite noun phrases*. Ph.D. thesis, University of Massachusetts, Amherst.

Kamp, H., & Reyle, U. (1993). *From discourse to logic: An introduction to modeltheoretic semantics, formal logic and discourse representation theory*. Dordrecht: Kluwer Academic Publishers.

Keller, W. (1988). Nested cooper storage: The proper treatment of quantification in ordinary noun phrases. In U. Reyle & C. Rohrer (Eds.), *Natural language parsing and linguistic theories* (pp. 432–447). Dordrecht: Reidel.

Kontinen, J., & Szymanik, J. (2008). A remark on collective quantification. *Journal of Logic, Language and Information*, *17*(2), 131–140.

Kontinen, J., & Szymanik, J. (2011). Characterizing definability of second-order generalized quantifiers. In: L.D. Beklemishev & R. de Queiroz (Eds.) *Proceedings of the 18th workshop on logic, language, information and computation, volume 6642 of lecture notes in computer science* (pp 187–200). Berlin. Springer. A journal version will appear in Journal of Computer and System Sciences WoLLIC 2011 special issue.

Krasikova, S. (2011). Definiteness in superlatives. In M. Aloni, V. Kimmelman, F. Roelofsen, G. Sassoon, K. Schulz, & M. Westera (Eds.), *Amsterdam colloquium on logic, language and meaning, volume 7218 of lecture notes in computer science* (pp. 411–420). Berlin: Springer.

Kratzer, A. (2007). On the plurality of verbs. In J. Dolling & T. Heyde-Zybatow (Eds.), *Event Structures in linguistic form and interpretation*. Berlin: Mouton de Gruyter.

Landman, F. (1998). Plurals and maximalization. In S. Rothstein (Ed.), *Events and grammar* (pp. 237–272). Dordrecht: Kluwer Academic Publishers.

Landman, F. (2000). *Events and plurality: The Jerusalem lectures*. Dordrecht: Kluwer Academic Publishers.

Link, G. (1983). The logical analysis of plurals and mass terms. In: R. Bauerle, C. Schwarze, & A. von Stechow (Eds.) *CSLI lecture notes, editor, meaning, use, and interpretation in language* (pp. 302-323). Berlin: de Gruyter.

May, R. (1985). *Logical form: Its structure and derivation*. Cambridge: MIT Press.

Montague, R. (1974). The proper treatment of quantification in ordinary English. In R. Thomason (Ed.), *Formal philosophy: Selected papers of Richard Montague* (pp. 247–270). New Haven: Yale University Press.

Mostowski, M., & Szymanik, J. (2012). Semantic bounds for everyday language. *Semiotica*, *188*(1–4), 363–372.

Neale, S. (1990). *Descriptions*. Cambridge: MIT Press.

Peters, S., & Westerståhl, D. (2006). *Quantifiers in language and logic*. Oxford: Oxford University Press.

Reimer, M. (1998). Quantification and context. *Linguistics and philosophy*, *21*(1), 95–115.

Robaldo, L. (2010a). Independent set readings and generalized quantifiers. *The Journal of Philosophical Logic*, *39*(1), 23–58.

Robaldo, L. (2010b). Interpretation and inference with maximal referential terms. *The Journal of Computer and System Sciences*, *76*(5), 373–388.

Robaldo, L. (2011). Distributivity, collectivity, and cumulativity in terms of (in)dependence and maximality. *The Journal of Logic, Language, and Information*, *20*(2), 233–271.

Sanford, A. J., Moxey, L. M., & Paterson, K. (1994). Psychological studies of quantifiers. *The Journal of Semantics*, *11*(3), 153–170.

Scha, R. (1981). Distributive, collective and cumulative quantification. In: J. Groenendijk, T. Janssen, & M. Stokhof (Eds.) *CSLI Lecture Notes, editor, formal methods in the study of language, part 2* (pp. 483–512). Mathematisch Centrum: Amsterdam.

Schein, B. (1993). *Plurals and events*. Cambridge, MA: MIT Press.

Schwarzschild, R. (1996). *Pluralities*. Dordrecht: Kluwer.

Sher, G. (1990). Ways of branching quantifiers. *Linguistics and Philosophy*, *13*, 393–422.

Sher, G. (1997). Partially-ordered (branching) generalized quantifiers: A general definition. *The Journal of Philosophical Logic*, *26*, 1–43.

Spaan, M. (1996). Parallel quantification. *Quantifiers, logic, and language* (Vol. 54, pp. 281–309). Stanford: CSLI Publications.

Stanley, J., & Szabò, Z. (2000). On quantifier domain restriction. *Mind and Language*, *15*, 219261.

Steedman, M. (2012). *Taking scope: The natural semantics of quantifiers*. Cambridge, MA: MIT Press.

Szabolcsi, A. (2013). Compositionality without word boundaries: (The) more and (the) most. In: *Proceedings of SALT* (Vol. 22).

Szymanik, J. (2009). *Quantifiers in TIME and SPACE. Computational complexity ofgeneralized quantifiers in natural language*. Ph.D. thesis, University of Amsterdam, Amsterdam.

Szymanik, J., & Zajenkowski, M. (2009). Comprehension of simple quantifiers empirical evaluation of a computational model. *Cognitive Science: A Multidisciplinary Journal*.

Szymanik, J. (2010). Computational complexity of polyadic lifts of generalized quantifiers in natural language. *Linguistics and Philosophy*, *33*, 215–250.

Szymanik, J., & Zajenkowski, M. (2013). Monotonicity has only a relative effect on the complexity of quantifier verification. In: F. Roelofsen, M. Aloni, & M. Franke (Eds.) *Proceedings of the 19th Amsterdam Colloquium* (pp. 219–225).

van Benthem, J. (1986). *Essays in logical semantics*. Dordrecht: Reidel.

van der Does, J. (1993). Sums and quantifiers. *Linguistics and Philosophy*, *16*, 509–550.
Winter, Y. (2001). *Flexibility principles in boolean semantics: Coordination, plurality, and scope in natural language*. Cambridge, MA: MIT Press.