RESOLVING WH/QUANTIFIER AMBIGUITIES:
A PSYCHOLINGUISTIC APPROACH

by

ASYA ACHIMOVA

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Psychology

written under the direction of

Julien Musolino, Ph.D.

and approved by

________________________________________

________________________________________

________________________________________

New Brunswick, New Jersey

January 2011
ABSTRACT OF THE THESIS

Resolving Wh/quantifier Ambiguities: a Psycholinguistic Approach

By ASYA ACHIMOVA

Thesis Director:
Julien Musolino, Ph.D.

The interaction of wh-words and quantifiers in questions has long been characterized as exhibiting a subject/object asymmetry, when a pair-list answer is available for object questions with subject quantifiers, like *Which book did every student read?* but not for subject questions with object quantifiers *Which student read every book?* (May 1985). Yet, the availability of pair-list answers to subject questions remains debatable and individual acceptability judgments reported in the literature vary significantly. We ran three experiments to test what factors proposed in the syntactic/semantic literature contribute to the availability of pair-list answers. The experiments confirmed the subject/object asymmetry. The data show that plurality of a *wh*-phrase does not make pair-list answers more readily available; pair-list answers are more likely to arise for questions with *each* than with *every*. We have also uncovered a group of individuals whose acceptability of pair-list answers is high both for subject and object questions and therefore does not show a subject/object asymmetry. We make an attempt to account for the behavior of that group by extending the analysis proposed in Beghelli (1997).
Acknowledgments

I would like to thank Julien Musolino, Viviane Deprez, Shigeto Kawahara, Kristen Syrett and Marta Suarez for their feedback and suggestions; Melanie Kelliher and Max Kon for help with experiments and data analysis; all the members of the Rutgers Psycholinguistics Lab, the audience of RuLing 2010 and NELS 41 for insightful comments and questions.
# Table of Contents

Abstract .............................................................................................................................................. ii  
Acknowledgements ...................................................................................................................... iii  
List of tables ....................................................................................................................................... v  
List of figures ...................................................................................................................................... vi  
0. Introduction ................................................................................................................................... 1  
1. Background ................................................................................................................................... 3  
   1.1 Theoretical approaches to wh/quantifier interaction .............................................................. 3  
   1.2 Methodology of data collection ................................................................................................. 8  
2. Experiments .................................................................................................................................. 10  
   2.1 Experiment 1. Who / which ...................................................................................................... 10  
   2.1.1 Method .................................................................................................................................. 10  
   2.1.2 Results and discussion ........................................................................................................... 13  
   2.2 Experiment 2. Each / every ........................................................................................................ 17  
   2.2.1 Method .................................................................................................................................. 17  
   2.2.2 Results and discussion ........................................................................................................... 18  
   2.3 Experiment 3. Which sg / Which pl ........................................................................................... 19  
   2.3.1 Method .................................................................................................................................. 19  
   2.3.2 Results and discussion ........................................................................................................... 20  
3. General discussion ......................................................................................................................... 22  
Bibliography ...................................................................................................................................... 25  
Tables ................................................................................................................................................. 26  
Figures .............................................................................................................................................. 30
Lists of tables

Table 1. Availability of pair-list answers for subject questions……………………………26
Table 2. Ordinal logistic regression. Experiment 1 Who / Which………………………….27
Table 3. Ordinal logistic regression. Experiment 2 Each / every…………………………..28
Table 4. Ordinal logistic regression. Experiment 3 Which sg. / Which pl………………29
List of figures

Figure 1. *Who did everyone see?* .......................................................................................30
Figure 2. *Who saw everyone?* ............................................................................................31
Figure 3. *Which boy kissed each girl?* ...............................................................................32
Figure 4. *Which student read every book?* ........................................................................33
Figure 5. Sample question .....................................................................................................34
Figure 6. Subject/object asymmetry.......................................................................................35
Figure 7. The participants assigned high ratings to appropriate answers.............................36
Figure 8. PL answers to object questions with subject quantifiers are accepted...............37
Figure 9. Inappropriate answers are rejected.........................................................................38
Figure 10. A group of people accept PL answers to subject question with object quantifiers .........................................................................................................................39
Figure 11. 30% of speakers do not show the subject/object asymmetry ...............................40
Figure 12. The type of a wh-phrase does not affect the ratings for PL answers ..................41
Figure 13. A correlation between ratings for PL answers to questions with *which* and *who* ....................................................................................................................................42
Figure 14. Strongly distributive quantifiers in the object position increase the availability of PL answers ...................................................................................................................43
Figure 15. Plurality of a wh-phrase does not increase the availability of PL answers.....44
0. Introduction

Wh/quantifier interaction is one of the central topics in the study of the syntax/semantics interface as it investigates the relations between logical form and surface form (Agüero-Bautista 2001). It has long been observed that questions containing universal quantifiers in argument positions may allow for at least two types of answers: a single answer (SA; 1a), and a pair-list one, (PL; 1b)\(^1\). Object questions with a subject quantifier, like (1), typically allow both answers. By contrast, the availability of a PL answer for subject questions with an object quantifier, as in (2), is more controversial.

(1) Who did everyone kiss?
(a) Everyone kissed John – SA
(b) Mary kissed John, Jane kissed Nick and Sarah kissed Michael – PL

(2) Who kissed everyone?
(a) Mary kissed everyone.
(b) ? Mary kissed John and Kelly kissed Nick.

This paper is aimed at resolving the apparent data disagreement concerning the availability of pair-list answers found in the literature (May 1985, Chierchia 1993, Beghelli 1997 among others). Establishing a stronger empirical base should help us determine a theoretical account that best fits the resulting empirical landscape. Moreover,

\(^1\) Endahl (1986), Chierchia (1993), Krifka (2003) distinguish between three readings of questions containing a quantifier: a narrow-scope reading (or a reading leading to a single answer), a pair-list and a functional reading. We will only consider two relevant types of readings in this paper, namely, those leading to single answers and pair-list answers.
since there is both growing evidence that controlled experimentations are useful tools in dealing with controversial cases (Kawahara in press, Sprouse & Almeida submitted, Syrett & Lidz to appear), as well as an on going methodological debate concerning the reliability of informal intuition judgments, our goal is to further contribute to this debate (Gibson & Fedorenko 2009, Culbertson & Gross 2009, Labov 1996, Schütze 1996, Phillips 2009, Sprouse 2009, Sprouse & Almeida submitted).

The paper is structured as follows: in section 1 we present some of the factors that have been claimed to affect the availability of PL answers, the empirical predictions that arise from theoretical accounts of wh/quantifier interaction. We show where these predictions agree and where they conflict, and how this shaped our research questions. This will lead to the presentation of our new empirical findings in section 2. In section 3 we discuss where our findings concur with distinct theoretical predictions and where they conflict. We conclude by suggesting possible interpretations of the conflicting data and where they could fit within the existing theoretical landscape.
1. Background

1.1 Theoretical approaches to wh/quantifier interaction

Several analyses have been proposed to account for what has long been regarded to be a characteristic subject/object asymmetry (May 1985, Cheirchia 1993, among others). The approaches not only differ in the mechanisms they use to account for the phenomenon but, crucially, in judgments about the facts themselves. One of the first comprehensive accounts of wh/quantifier interaction at LF was developed by May (1985). May predicts that while object questions with subject quantifiers (1) generally allow both a SA and a PL answer, subject questions with object quantifiers like (2), lack a pair-list reading. For May (1985), the distinction is syntactic and relates to other subject/object asymmetries claimed to exist in the literature, such as for instance the well-known that-trace effect (Pesetsky 1982 among others). Syntactic structure is taken to be a central factor governing the availability of PL answers. For a pair-list answer to be available, a quantifier must take scope over the wh-expression at LF, which obtains when the quantifier and a wh-phrase form a Sigma sequence after the quantifier raises and adjoins to IP. The process is constrained by the Path Containment Condition (Pesetsky 1982). Pair-lists are available for subject quantifiers interacting with object wh-phrases because the paths of both nest, as shown in Figure 1.

-------------------------------
*Insert Figure 1 about here.*
-------------------------------
For questions with object quantifiers, PL is not available since the paths resulting from quantifier raising and wh-movement cross, as illustrated in Figure 2; the quantifier must stay lower in the tree and no PL is possible.

Chierchia (1993), though developing a different approach to wh/quantifier interaction, still views structural asymmetry as a fundamental phenomenon. According to Chierchia (1993), PL answers result from the binding by a quantifier of a functional variable left by the movement of the question-term. This quantificational binding, like that of a pronoun, is constrained by Weak Crossover, which is also structurally determined. A configuration, when an object quantifier binds a variable left by a subject wh-term after its raising to Spec IP, results in WCO, and a pair-list is ruled out. This account predicts no PL answer to be available for subject questions with object quantifiers, like (2).

To account for some apparent exceptions in examples such as (3), Chierchia further suggests that semantically plural wh-phrases allow list answers in a way similar to how plural pronouns escape the WCO constraints in examples like (4).

(3) *Who put everything on the platter? PL ok.

(4) Their mothers like every boy in the class.

A structure like (4) should typically result in a WCO yielding ungrammaticality, like (5). It is the plurality of a pronoun is (4) that allows avoiding WCO effect in (4).

(5) *His, mother loves everyone,.
For Chierchia (1993), pair-list answers can be available for questions with *who* interacting with an object quantifier, because *who* can be semantically plural, but not for question with singular *which*. So subject/object asymmetry obtains only for a subset of questions and does not arise for subject questions with question-words like *who* that allow a plural reading.

Beghelli (1997) in his analysis of pair-list answers develops an idea initially suggested by Williams (1988) that the availability of pair-list answers depends on the type of the quantifier the question-term is interacting with. Strongly distributive quantifiers like *each* are always raised to the specifier of DistP which is higher than IP, so always allow inverse scope of an object over a subject. *Each* in (6) raises to DistP and can bind the variables introduced by the wh-phrase, so a PL is available (Figure 3).

(6) Which girl kissed each boy? PL ok.

Weakly distributive quantifiers like *every* on the other hand, introduce a set variable that, when bound by the wh-operator, constrains the inverse scope of an object over a subject. As a result the subject/object asymmetry only obtains with weakly distributive and non-distributive quantifiers.

Yet another different approach is proposed by Agüero-Bautista (2001). The availability of pair-list answers is not so much restricted by the structure but more importantly by discourse-related properties of the wh-phrase. While *which* is lexically

---

2 Beghelli (1997) relies on Szabolcsi’s (1997) classification of quantifiers and treats both *every* and *each* as being strongly distributive. Yet, when *every* interacts with a wh-term, the set variable introduced by the universal quantifier is bound, resulting in the loss of strong distributivity.
presuppositional, *who* is not. The presuppositionality becomes important because according to Agüero-Bautista, presuppositional wh-phrases cannot reconstruct into the theta-position they originated from. In Agüero-Bautista’s analysis, reconstruction below a quantifier is a necessary condition for a pair-list answer to be available, and this explains why a PL is unavailable in (7).

\[(7) \text{Which student read every book? * PL}\]

A wh-phrase can reconstruct in Spec IP but not in Spec vP. Only the latter is a position below the quantifier, so a PL is impossible (Figure 4).

Insert Figure 4 about here.

In contrast, *who* can sometimes reconstruct, because the presuppositional status of *who* is determined by discourse, not by the lexicon. Similar to Chierchia’s (1993) analysis, Agüero-Bautista predicts that a PL answer is available for a subject *who* question interacting with object *every*. Moreover, in accord with Beghelli’s view (1997), Agüero-Bautista claims that *each* moves through quantifier raising to a position higher than *every*, which allows it to take scope over the trace of the wh reconstructed in Spec IP. This predicts that with *each* no subject/object asymmetry will arise.

Using examples from Spanish he also argues that it is not the plurality of a *wh*-word that makes a PL answer possible, which goes against Chierchia’s (1993) approach. Agüero-Bautista concludes that *wh*/quantifier interactions do not always give rise to a subject-object asymmetry; instead the asymmetry is a phenomenon restricted only to a subset of wh-phrases and quantifiers, and only describes the cases that involve
presuppositional or definite interrogative determiners (Agüero-Bautista 2001: 54). Presuppositional wh-phrases like *which* cannot reconstruct below the quantifier and a pair-list answer is not available for them. Since *who* is not presuppositional, it can yield a pair-list answer when interacting with a quantifier.

Table 1 summarizes the empirical predictions of the different accounts we just briefly reviewed. All the accounts predict a PL answer not to be available for questions where a presuppositional wh-phrase interacts with an object *every*. Chierchia (1993) predicts pair-list answers to be possible for questions with a wh-phrase that can be plural, like *who*. This entails that PL answers should also be possible for questions in which a plural which-phrase interacts with an object quantifier. The latter structure is predicted to lack a PL reading in the approach of Agüero-Bautista (2001). PL answers are predicted to be available for subject questions with object each (Chierchia 1993, Agüero-Bautista 2001).

| Insert Table 1 about here |

Though there is disagreement about the predictions and the data, as the table shows, there is also a case of complete convergence. We have chosen the case of *which* subject + object *every* as a baseline for our experiment as the approaches discussed above make the same predictions, namely that a PL answer is impossible in this case, although for different reasons. We tested this base line case against object questions in which PLs are assumed to be always possible and against subject and object questions in which we manipulated the semantic and morphological plurality vs. singularity of the question
term, the presuppositional vs. non presuppositional nature of the wh-term and, finally, the strong vs. weak distributive character of the quantifier.

Beyond the original syntactic position of the question and quantifier terms, a number of additional factors have been claimed to play a role: the type of the question-word (Chierchia 1993, Agüero-Bautista 2001), the nature of the interacting quantifier (Williams 1988, Agüero-Bautista 2001), their number and their presuppositional status (Agüero-Bautista 2001). To assess the contribution of these factors, we conducted a series of psycholinguistic experiments that directly manipulated some of the variables described above. Experimental method also allowed us to get objective data to resolve the controversies that exist in the literature.

1.2 Methodology of data collection


Methodology of obtaining informal judgments has received critique in the works of Spencer (1973), Schutze (1996), Wasow and Arnold (2005), Gibson and Fedorenko (2010), Dabrowska (2010) and others. The following properties associated with informal judgments methodology are claimed to be problematic: small sample sizes (usually the researcher and a few other speakers), bias (authors report their own judgments), low generalizability of the observed phenomenon and others (Kawahara, in press).

Despite the critique of the traditional linguistic methodology it has been shown that many linguistic effects are robust and hold with a low number of judgments and
participants (Sprouse & Almeida submitted). It is claimed that eliciting judgments in informal settings is generally efficient for theory construction, and widely accepted linguistic generalizations based on professional intuitive judgments overall turn out to hold when tested under more formal experimental designs (Phillips 2009; Sprouse 2009). Sprouse and Almeida (submitted) performed a series of experiments looking at island effects (Ross 1967). Using resampling simulations they show that with a single judgment per participant given the smallest sample size of 5 there is a 70-75% detection rate for the following types of island effects: Whether, Subject and Adjunct islands. The detection rates rise to over 90% for these types of islands when four judgments per participant are used. The results show that even using a sample size of just five participants and one judgment per speaker would reveal the predicted effects. Moreover, the study shows that participants never report the effect in the opposite direction, in other words speakers never prefer a sentence without island violations to those containing island violations. The authors conclude that informal judgments reported in the literature are in fact quite robust and there is no real evidence that informal acceptability judgments are unreliable.

Yet, when there is disagreement in judgments as to whether particular sentence types are acceptable or not, it becomes difficult to decide which judgments, amongst the ones reported in the literature, reflect data accurately (Grudzinska 2008). Using experimental method offers a new possibility of resolving disagreement or, at the very least, obtaining a more detailed picture of the phenomenon. The distribution of judgments elicited from a large number of speakers as we later show can reveal more variation in responses than predicted by theoretical accounts (May 1985, Beghelli 1997, Chierchia 1993, Agüero-Bautista 2001) and therefore provide motivation for theory elaboration.
2. Experiments

In an attempt to resolve the disagreement about the data and find out what factors affect the availability of pair-list answers, we ran three experiments using judgment tasks. Undergraduate students naïve to linguistic theories participated in the experiments to ensure that there is no effect of bias (Gibson & Fedorenko 2010; for opposite views see Sprouse & Almeida submitted). The following parameters of the stimuli were we controlled for: keeping the questions/answers as close to those discussed in the literature as possible, creating minimal subject/object pairs, having a set of controls to ensure that participants understood the task, could assign low/middle/high ratings when required, and accepted pair-list answers when they were appropriate.

Participants (n=95, all native speakers of English) were asked to rate answers to questions containing a wh-expression and a quantifier on a 1-7 scale. We used a scale instead of a binomial choice (i.e., Yes/No) because of the variation in the judgments originally found in the syntactic literature. A scale can show not only mean ratings assigned by participants, but, crucially, the degree of variation that exists in judgments.

2.1 Experiment 1 Who / which

2.1.1 Method

*Design.* Experiment 1 was designed to investigate whether the subject/object asymmetry (May 1985) can be confirmed experimentally when tested with a large sample size of participants (n=33) and four tokens per condition. Second, we tested whether the nature of the wh-expression involved (*which* vs. *who*) (Chierchia 1993, Agüero-Bautista 2001) affects ratings for pair-list answers. In this experiment, we manipulated answer
type; grammatical position of the wh-term and the type of a wh-term used (lexically vs. discourse presuppositional wh-phrases as well as possibly plural vs. singular wh-phrases). Thus, experiment 1 had a 2x2x2 design (2 (grammatical position: subject vs. object wh) x 2 (answer type: single vs. pair-list) x 2 (wh-type: who vs. which) in which all three factors were treated as within-subjects variables. Crossing of these factors results in eight conditions:

1) *Who kissed every boy?* Single answer;
2) *Who did every boy kiss?* Single answer;
3) *Which girl kissed every boy?* Single answer;
4) *Which girl did every boy kiss?* Single answer;
5) *Who kissed every boy?* Pair-list answer;
6) *Who did every boy kiss?* Pair-list answer;
7) *Which girl kissed every boy?* Pair-list answer;
8) *Which girl did every boy kiss?* Pair-list answer.

*Task.* Participants were asked to rate 32 critical items (8 conditions, 4 lexical sets per condition) and 60 control/filler statements, the latter included answers to questions with wh-words only, quantifiers only, questions with clearly acceptable or unacceptable answers, as well as questions with pragmatically odd answers. Four lists were created in which order of items was randomized, and in each experiment, participants were randomly assigned to a list.

*Software.* We used Survey Monkey software to run the experiments. Analyses were performed using PASW 17.
Statistical analysis. Significance testing is often performed by means of ANOVA which is claimed to be robust to various violations of its basic assumptions. Yet, a parametric test, such as ANOVA, might not be accurate when dealing with categorical data, which is not normally distributed. Applying ANOVA in that case may easily create misleading results (Jaeger 2008). Ordinal logistic regression is an alternative analysis which can be applied to data obtained from experiments which involved using a scale. Logistic regression models do not assume linearity, nor do they require homogeneity of variance which, as we are going to show below, is particularly important in our case. Both analyses were performed and the results turned out to be similar, so we use the results from logistic regression in the main body of the text. In cases when we found discrepancies between the two tests we adhere to logistic regression results.

Participants. Thirty-three native speakers of English participated in the experiment. All participants were Rutgers undergraduate students. They received course credit for participation.

Procedure. Each experiment started with the presentation of three trial stimuli, which showed possible (8), impossible (9) and ‘intermediate’ answers (10).

(8) Did you see an elephant at the zoo?
Yes, I did.

(9) Where are you from?
I don’t really like chocolate.

(10) What kind of music do you prefer?
Oh, I like music!
Participants then took the main test which usually lasted for 15-20 minutes. There was no time pressure on the participants: they could spend as much time as they needed on each question but they were not allowed to go back and change their responses. All responses were recorded automatically by the software.

Each trial represented a pair of a question and an answer to that question. The task was to determine whether that answer was possible on a 1 - 7 scale (Figure 5).

2.1.2 Results and discussion

To find out whether certain factors affect availability of pair-list answers, we compared acceptability ratings that participants assigned in experimental conditions using ordinal logistic regression. Table 2 presents results of ordinal logistic regression. Results that are significant are highlighted in light gray.

The analysis revealed a significant effect of the answer type (p < 0.01), and the effect of the grammatical position of a wh (p < 0.01). For subject questions (2), pair-list answers were significantly less acceptable than for object questions (1) (p< 0.01). This reflects the classic subject/object asymmetry described by May (1985) which predicts pair-list answers to be available for object questions (high ratings) and not to be available for subject questions (lower ratings). Figure 6 shows that for questions with both who and
which SA ratings are high, it means that the speakers accepted SA as predicted. It also shows that the participants used the scale correctly. The bars representing PL answer ratings are lower for subject questions with an object quantifier, than for an object wh-phrase interacting with a subject quantifier. This pattern holds both for questions with who and which.

Singular presuppositional wh-phrases, like which, are predicted to lack a PL reading, a conclusion shared by all theoretical accounts reviewed in this paper (May 1985, Chierchia 1993, Beghelli 1997, Agüero-Bautista 2001). At the same time, PL answer ratings for subject questions with object quantifiers are relatively high for an answer that is supposed to be rejected by grammar. PL answer ratings for subject questions with an object quantifier are also significantly higher than ratings for unacceptable answers in the control conditions (p < 0.01, Kruskal-Wallis test).

The analysis uncovered that at least 30% of the participants consistently assigned a rating of 6 or 7 to pair-list answers to subject questions with object quantifiers, predicted to lack a pair-list reading (May 1985, Beghelli 1997). To ensure that the observed pattern of responses does not result from speakers not understanding/ not paying attention to the experimental task we compare their performance in experimental conditions to the one in control conditions. Figure 7 shows that speakers can assign a high rating to an appropriate answer as expected.
Pair-list answers are expected to be acceptable for object questions with a subject quantifier *every*, Figure 8 illustrates that speakers behave exactly as predicted and assign high ratings to PL answers.

When expected to reject an answer, the speakers assigned a low rating (Figure 9, left plot). They were also sensitive to intermediate levels of ‘appropriateness’ of an answer and can use the middle of the scale when necessary (Figure 9, right plot).

If the predictions of May (1985) are correct and our grammar does not generate a pair-list answer as a response to a subject wh-questions with an object *every*, we are supposed to see in Figure 10 that speakers reject a pair-list answer. What we actually see here is a cluster in the left part of both graphs and, quite unexpectedly, a cluster in the right part of the graphs, reflecting the fact that a considerable number of participants assigned high ratings to pair-list answers to subject questions with an object quantifier.
Figure 11 plots the acceptability ratings for subject question PL answers against the ratings for object question PL answers. Dots falling on or close to the identity line indicate speakers whose responses show no subject/object asymmetry, contrary to the predictions (May 1985, Chierchia 1993 among others). At the same time there are no data points below the line, showing that PL answer ratings for questions with object quantifiers never receive higher ratings than the ones with subject quantifiers, correctly predicted by the theoretical accounts (May 1985, Beghelli 1997 among others).

The results of experiment 1 revealed that who did not make pair-list answers more available than which (p = 0.347), contra the predictions that the plurality of who (Chierchia 1993), or the ability of who to be reconstructed (Agüero-Bautista 2001) should make pair-list answers possible with who but not with which (Figure 12).

To support the finding that who and which are not significantly different in their ability to give rise to pair-list answers, we plotted the ratings for PL answers to questions with who in subject position interacting with every against the PL ratings for subject which interacting with the same object quantifier. Figure 13 shows a strong correlation between ‘who’ and ‘which’ ratings (r = 0.85, p < 0.01). This suggests that people who
accepted pair-list answers with *who* also accepted pair-list answers to questions with *which*, contrary to the theoretical predictions (Chierchia 1993).

*Insert Figure 13 about here.*

Results of experiment 1 show that, as predicted, subject/object asymmetry holds overall for the majority of speakers. At the same time for at least 30% of the participants, ratings for PL answers to subject and object questions are not significantly different from each other. The possible plurality of *who* compared to *which* sg does not increase acceptance rates for PL answers to subject questions with object quantifiers.

### 2.2 Experiment 2 Each / every

#### 2.2.1 Method

*Design.* Experiment 2 examined the contribution of the quantifiers (*every* vs. *each*) to the availability of PL answers (Williams 1988, Beghelli 1997, Szabolcschi 1997, Agüero-Bautista 2001). In a 2x2x2 design the following factors were tested: the type of answer; grammatical position of a wh-term and the type of the quantifier (*each* vs. *every*) while the type of wh was constant in this experiment and fixed to *which* sg value. The following eight conditions were tested with four lexical sets per condition:

1) **Which girl kissed every boy?** Single answer;

2) **Which girl did every boy kiss?** Single answer;

3) **Which girl kissed each boy?** Single answer;

4) **Which girl did each boy kiss?** Single answer;
5) *Which girl kissed every boy?* Pair-list answer;

6) *Which girl did every boy kiss?* Pair-list answer;

7) *Which girl kissed each boy?* Pair-list answer;

8) *Which girl did each boy kiss?* Pair-list answer.

**Participants.** Twenty-nine native speakers of English participated in the experiment. All participants were Rutgers undergraduate students. They received course credit for participation.

**Procedure.** Experimental procedure was similar to experiment 1 with the same number of critical and control items. The same software was used to display the stimuli and collect responses.

### 2.2.2 Results and discussion

Summary of the results of experiment 2 are presented in Table 3.

---------------------------------

Insert Table 3 about here.

-----------------------------------

The analysis revealed a significant effect of the answer type (p < 0.01), the position of a wh-phrase (p < 0.01), and type of a quantifier (p < 0.01). The results confirmed subject/object asymmetry with object questions PLs receiving higher ratings than subject questions PL answers, confirming predictions of the theoretical accounts (May 1985 among others) and the findings of experiment 1 (p < 0.01). Pair-list answers are more readily available for subject questions with *each* than with *every* (p < 0.01), as predicted by Beghelli (1997), Agüero-Bautista (2001) (Figure 14). This suggests that it is
the distributivity of a quantifier that affects the availability of PL answers to subject questions.

Insert Figure 14 about here.

2.3 Experiment 3 Which sg / Which pl

2.3.1 Method

Design. Experiment 3 tested the role played by the plurality of the wh-phrase (\textit{which}-singular (12) vs. \textit{which}-plural (13)) (Chierchia 1993, Agüero-Bautista 2001).

\begin{enumerate}
\item (12) \textit{Which student read every book on the list?}
\item (13) \textit{Which students read every book on the list?}
\end{enumerate}

Three factors were manipulated in the experiment: wh-position (subject vs. object), the type of answer (SA vs. PL answer) and the grammatical number of a wh-phrase (singular wh-phrases vs. plural wh-phrases). Eight conditions were represented by 4 lexical sets each:

\begin{enumerate}
\item \textit{Which girl kissed every boy?} Single answer;
\item \textit{Which girl did every boy kiss?} Single answer;
\item \textit{Which girls kissed every boy?} Single answer;
\item \textit{Which girls did every boy kiss?} Single answer;
\item \textit{Which girl kissed every boy?} Pair-list answer;
\item \textit{Which girl did every boy kiss?} Pair-list answer;
\item \textit{Which girls kissed every boy?} Pair-list answer;
\item \textit{Which girls did every boy kiss?} Pair-list answer;
\end{enumerate}
8) Which girls did every boy kiss? Pair-list answer.

Participants. Thirty-three native speakers of English participated in the experiments. All participants were Rutgers undergraduate students. They received course credit for participation.

Procedure. Experimental procedure was the same as in experiments 1 and 2. The same software was used to display the stimuli and collect responses.

2.3.2 Results and discussion

The analysis showed the significant effect of the answer type with single answers receiving higher ratings overall than pair-list answers (p < 0.05); and the significant effect of wh-position (p < 0.01). The interactions of wh-position and answer type, as well as the interaction of answer type and wh-number are significant (p < 0.01). The latter interaction is driven by single answer ratings: they are lower to questions with a plural wh-phrase than to questions with a singular wh-phrase. Summary of the results are presented in Table 4.

Insert Table 4 about here.

The analysis revealed no significant effect of wh-number overall across conditions (p = 0.292). The difference between questions with singular which (12) and plural which (13) in their ability to yield pair-list answers is not significant either (p = 0.367) (Figure 15). This finding goes against Chierchia’s (1993) hypothesis and confirms Agüero-Bautista (2001) generalization based on data from Spanish that the plurality of a wh-
phrase does not affect the availability of PL answers to subject questions with object quantifiers.

----------------------------------

*Insert Figure 15 about here.*

----------------------------------

Results of experiment 3 replicate the findings of experiments 1 and 2 and confirm the prediction that PL answers are more acceptable for object questions with subject quantifiers than for subject questions with object quantifiers (May, 1985 among others). Contra Chierchia (1993) and confirming Agüero-Bautista (2001), plurality of a wh-phrase does not make PL answers more readily available for subject questions interacting with an object *every.*
3. General discussion

Overall, the subject/object asymmetry holds for questions with the universal quantifier *every*: PL answers to subject questions with object quantifiers (2) receive on average lower ratings than questions like (1). This finding was replicated in all the three experiments. Yet, the ratings for PL answers to subject questions with object *every* are high for a type of answers that is supposed to be rejected by grammar. It is not only higher than unacceptable controls but also above the mean rating for controls showing intermediate levels of appropriateness. This observation raises a question whether the constraints that regulate the availability of PL answers are purely grammatical. At the same time, there are no speakers who assign higher ratings to subject question PL answers than to object question PL answers, as correctly predicted by May (1985). Given the data we observed, we assume that the access to PL readings is regulated by a complicated interaction of grammatical and discourse factors.

We observed more variation in the responses than predicted by structural accounts for subject questions with object quantifiers; the asymmetry does not hold for at least 30% of the speakers in our experiments. This finding supports the claim that grammatical factors are not sufficient to account for the conditions that license PL answers. Given the grammatical account only, we would have to assume that those speakers’ grammar is different from the other 70% of the speakers. On the other hand, if we develop a model where grammatical constraints interact with discourse factors we might be able to account for the variability in the responses and preserve the current analyses of wh/quantifier interaction.
Morpho-syntactic features associated with the plurality of a wh-phrase do not increase acceptability ratings for pair-list answers to subject questions (contra Chierchia 1993). This conclusion also shows that not all the intuitions reported in the literature can be verified experimentally. What seems to be more relevant to the availability of PL answers is the distributive status of the interacting quantifier. Confirming predictions of Beghelli (1997) we show that distributive quantifiers give rise to significantly higher acceptance rates of pair-list answers to subject questions.

Discourse factors did not show a clear effect on the availability of pair-list answers. PL answers to questions with subject who were accepted by some speakers and rejected by others, which can be explained by the fact that without context speakers differ in the level of presuppositionality they attribute to who. If the question term was treated as presuppositional, on other words close to which, reconstruction into a theta-position was impossible resulting in the rejection of a PL answer. If on the other hand, who was treated as a non-presuppositional wh-phrase, nothing constrained reconstruction and a PL was possible. Such an interpretation of the results could possibly explain variation in ratings for PL answers to questions with subject who. At the same time it is unclear how this analysis could be extended to explain the variation in responses to questions with which. Being lexically presuppositional, the latter is predicted to be unable to reconstruct into a theta-position and no PL should be available. This is clearly not the case for at least 30% of our participants.

It seems though that Beghelli’s account of quantifier distributivity can be extended to explain the data we observed. Our suggestion at this point is stipulative. If some speakers treat every as a strongly distributive quantifier, which it is, and it remains
strongly distributive even when interacting with a wh-term, then it would be possible to say that those speakers who accepted a PL, treat every as each, which is known to facilitate the access to a PL reading.

Overall, our findings indicate that the data surrounding wh-quantifier interaction are more complex than predicted by earlier syntactic theories with at least 30% of speakers not exhibiting subject/object asymmetry. Moreover, the nuanced array of facts that we uncovered - sometimes at odds with more recent accounts - underscores the role of controlled experimentation as an important tool for theory construction. More importantly, from a theoretical perspective our results suggest that either:

(a) There are two grammars in the population: one with a subject/object asymmetry (the majority of speakers) and one without (at least 30% of speakers); or

(b) There is in fact only one grammar, corresponding to the group accepting pair-list readings with object quantifiers, but, crucially, that it does not contain a subject/object asymmetry. In this case, the low ratings for the group that does not accept pair-list readings with object quantifiers could be attributed to other, extragrammatical factors. Either way, this conclusion represents a significant departure from current understanding and has a range of implications for theory, methodology, and acquisition.

The results show the benefits of running formal experiments in cases that have been claimed controversial. Obtaining responses from a larger group of speakers allowed us to explain the variation initially reported in the literature. Disagreement in reported judgments and predictions reflects actual variability that we observe among speakers, and therefore should be treated not as contradictory evidence but instead as a fuller reflection of wh/quantifier interaction.
Bibliography


Tables

Table 1. Availability of pair-list answers for subject questions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Who kissed every girl?</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Which boy kissed every girl?</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Which boys kissed every girl?</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Which boy kissed each girl?</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

The table shows the predictions of different accounts concerning the availability of PL answers to subject questions with object quantifiers. All the accounts predict a PL answer not to be available for a singular presuppositional wh-phrase which + NP. Other cases cause disagreement between the authors.
Table 2. Ordinal logistic regression. Experiment 1 Who/Which

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of answer (single vs. pair-list)</td>
<td>2.116</td>
<td>0.688</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of wh (who vs. which)</td>
<td>-0.228</td>
<td>0.531</td>
<td>0.668</td>
</tr>
<tr>
<td>Grammatical position of wh (subject vs. object)</td>
<td>-1.512</td>
<td>0.538</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of answer x grammatical position (subject vs. object)</td>
<td>2.144</td>
<td>0.790</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of answer x type of wh</td>
<td>-0.620</td>
<td>0.784</td>
<td>0.429</td>
</tr>
<tr>
<td>Grammatical position x type of wh</td>
<td>0.171</td>
<td>0.684</td>
<td>0.802</td>
</tr>
</tbody>
</table>

The table summarizes the results of regression analysis of data from Experiment 1. Significant effects and interactions are highlighted in light gray.
Table 3. Ordinal logistic regression. Experiment 2 Each/every

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of answer (single vs. pair-list)</td>
<td>2.376</td>
<td>0.599</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of quantifier (each vs. every)</td>
<td>1.973</td>
<td>0.553</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Grammatical position of wh (subject vs. object)</td>
<td>-1.709</td>
<td>0.489</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of answer x grammatical position (subject vs. object)</td>
<td>3.452</td>
<td>1.219</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of answer x type of quantifier</td>
<td>-3.796</td>
<td>0.822</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Grammatical position x type of quantifier</td>
<td>-0.294</td>
<td>0.726</td>
<td>0.685</td>
</tr>
</tbody>
</table>

The table summarizes the results of regression analysis of data from Experiment 2. Significant effects and interactions are highlighted in light gray.
Table 4. Ordinal logistic regression. Experiment 3 Which sg./Which pl.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter estimate</th>
<th>Standard error</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of answer (single vs. pair-list)</td>
<td>1.109</td>
<td>0.494</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Plurality of wh (which sg. vs. which pl.)</td>
<td>-0.482</td>
<td>0.458</td>
<td>0.292</td>
</tr>
<tr>
<td>Grammatical position of wh (subject vs. object)</td>
<td>-2.883</td>
<td>0.501</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Type of answer x grammatical position</td>
<td>4.878</td>
<td>0.965</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>(subject vs. object)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type of answer x plurality of wh</td>
<td>2.477</td>
<td>0.940</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Grammatical position x plurality of wh</td>
<td>0.571</td>
<td>0.632</td>
<td>0.367</td>
</tr>
</tbody>
</table>

The table summarizes the results of regression analysis of data from Experiment 3. Significant effects and interactions are highlighted in light gray.
For a PL answer to arise, the quantifier needs to take scope over the wh-phrase, so the quantifier adjoins to IP to yield a Sigma sequence. PCC is satisfied, therefore movement is possible. Members of the Sigma sequence can take scope over each other, so a PL answer is available.
Figure 2. *Who saw everyone?*

Adjunction to IP is constrained by PCC. The quantifier and the wh-phrase cannot form a Sigma sequence at LF, so the quantifier is not able to take scope over the wh-term, resulting in the impossibility of a PL answer.
Figure 3. *Which boy kissed each girl?*

Expressions in **bold** show possible reconstruction sites. In order to obtain a PL reading, the wh-phrase needs to reconstruct below the quantifier. A PL answer is available for questions where a wh-term interacts with an object *each*.
Figure 4. *Which student read every book?*

Expressions **in bold** show possible reconstruction sites. In order to obtain a PL reading, the wh-phrase needs to reconstruct below the quantifier. The lowest possible reconstruction site in this structure is Spec IP, which is not low enough to occur below the quantifier, so a PL reading is impossible.
Figure 5. Sample question

(11) Which driver took everybody home last night?

*Tom took Ms. Franko, Bob took Ms. Dombovski, and Jack took Mr. Perkins.*

Is that a possible answer? 1(definitely no) 2 3 4 5 6 7(definitely yes)

On each trial a participant saw a pair of a question and an answer to that question. The task was to rate the possibility of that answer on a 1 – 7 scale.
Figure 6. Subject/object asymmetry

Light gray bars represent questions containing a subject wh-phrase and an object quantifier. Dark gray bars represent questions with an object wh-phrase and a subject quantifier. Ratings for SA are high both for questions with who and which, as predicted. PL answers for subject questions with an object quantifier receive lower ratings than PL answers to object wh-questions with a subject quantifier. This constitutes the subject/object asymmetry.
Figure 7. The participants assigned high ratings to appropriate answers

The histograms show the distribution of ratings for control items. As expected, the participants assigned a high rating to an answer that was appropriate (both panels). The right panel also shows that the participants accepted PL answers when those were available (multiple wh-questions).
Figure 8. PL answers to object questions with subject quantifiers are accepted

The histograms represent the distribution of ratings for PL answers to object questions with subject quantifiers. The bars are clustered in the right end of the scale showing that the participants accepted PL answers, exactly as expected.
Figure 9. Inappropriate answers are rejected

The histograms plot the distribution of ratings for answers that are predicted to be rejected (left panel) and answers that show intermediate levels of appropriateness. The participants’ ratings show that they can use scale as intended and understand the task.
Figure 10. A group of people accept PL answers to subject question with object quantifiers.

The histograms show the distribution of ratings for PL answers to subject questions with object quantifiers. PL answers to questions with *which*, are predicted not to be available. They are rejected by a number of speakers; at the same time there is a cluster of bars on the right, showing that a certain number of speakers could access the supposedly unavailable reading. The right panel shows a similar pattern, which confirms that questions with *who* and *which* are not different in their ability to give rise to PL answers, contrary to the theoretical predictions.
Figure 11. 30% of speakers do not show the subject/object asymmetry

The graph plots the ratings for PL answer to subject questions with object quantifiers against the ratings for PL answers to object questions. All data points are clustered above the line, which shows that no participant rates PL answers to subject questions higher than those to object questions. At the same time, a cluster of data points in the top right corner are the speakers who do not show a subject/object asymmetry.
Figure 12. The type of a wh-phrase does not affect the ratings for PL answers

<table>
<thead>
<tr>
<th>Type of wh</th>
<th>Subject questions with object every</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean rating</td>
<td></td>
</tr>
<tr>
<td>Who</td>
<td>4</td>
</tr>
<tr>
<td>Which</td>
<td>5</td>
</tr>
</tbody>
</table>

Mean ratings for pair-list answers to subject questions with an object *every*. There is no significant difference between the two groups of ratings.
Figure 13. A correlation between ratings for PL answers to questions with *which* and *who*

This graph shows a correlation between the ratings for PL answers to subject questions with *who* and *which*. Contrary to the predictions, the participants who accept PL answers to questions with *who* interacting with an object quantifier, also accept PL answers to questions with *which*. 
Figure 14. Strongly distributive quantifiers in the object position increase the availability of PL answers

The graph plots mean ratings for SA and PL answers to questions with strongly distributive object quantifier *each* and a weakly distributive quantifier *every*. Ratings are similar for SA. PL answers are more readily available for questions with *each*.
Figure 15. Plurality of a wh-phrase does not increase the availability of PL answers

Plurality of a wh-phrase does not increase the availability of PL answers, there is no significant difference between the scores for singular and plural wh-questions.