
How do 5 year olds understand questions? Differences in languages across Europe

Journal Title
XX(X):2–35
© The Author(s) 2015
Reprints and permission:
sagepub.co.uk/journalsPermissions.nav
DOI: 10.1177/ToBeAssigned
www.sagepub.com/


Uli Sauerland¹, Kleanthes K. Grohmann², Maria Teresa Guasti³, Darinka Anđjelković⁴, Reili Argus⁵, Sharon Armon-Lotem⁶, Fabrizio Arosio³, Larisa Avram⁷, Joao Costa⁸, Ineta Dabašinskienė⁹, Kristine de Lopez¹⁰, Daniela Gatt¹¹, Helen Grech¹¹, Ewa Haman¹², Angeliek van Hout¹³, Gordana Hrzica¹⁴, Judith Kainhofer¹⁵, Laura Kamandulytė-Merfeldienė⁹, Sari Kunnari¹⁶, Melita Kovačević¹⁴, Jelena Kuvac Kraljević¹⁴, Katarzyna Lipowska¹⁷, Sandrine Mejias¹⁸, Maša Popović⁴, Jurate Ruzaitė⁹, Maja Savić⁴, Anca Sevcenco⁷, Spyridoula Varlokosta¹⁹, Marina Varnava², and Kazuko Yatsushiro²⁰

Abstract

The comprehension of constituent questions is an important topic for language acquisition research and for applications in the diagnosis of language impairment. We present the results of a study investigating the comprehension of different types of questions by five year old, typically developing children across 19 European countries, 18 different languages, and 7 language (sub-)families. We studied the effects of two factors on question formation: a) whether the question contains a simple interrogative word like *'who'* or a complex one like *'which princess'*, and b) whether the question word was related to the sentential subject or object position of the verb. We found that there is considerable variation among languages, but the two factors mentioned consistently affect children's performance. The cross-linguistic variation shows that three linguistic factors facilitate children's understanding of questions: having overt case morphology, having a single lexical item for both *'who'* and *'which'*, and the use of synthetic verbal forms.

Keywords

questions, case, agreement, cross-linguistic, syntax, comprehension, wh-phrases

Introduction

Research in language acquisition is most frequently conducted with a focus on one or two particular languages. Such research has led to many interesting findings, including differences between languages. But studies designed only for one or two languages are often not the best way to address questions of how specific properties of the target language affect the acquisition process. The reason is that such studies often need to be adjusted in non-trivial ways before they can be carried out for a different language which makes it impossible to compare the results across studies. Recently a number of works has appeared that compare a larger sample of languages (Armon-Lotem et al., 2015; Katsos et al., 2012; Tribushinina et al., 2013; Varlokosta, Belletti, et al., 2015; Xanthos et al., 2011) within a single study. These studies have addressed passives, quantification, pronouns, inflection, and adjectives. The study we report is the first to target constituent questions for a larger sample of languages.

The understanding of constituent questions represents an important area of investigation in language acquisition studies (Thornton 1990; van der Lely and Battell 2003; van der Lely and Pinker 2014, and others). The interest of this area is both theoretical and practical. The theoretical interest derives from the fact that questions are distinguished from declarative sentences in many languages by word order. Most prominently, the questioned constituent occupies an initial position in English and most other European languages. Especially interesting is that, as a consequence, the order of subject and object in a question can deviate from the canonical order of the two arguments. Consider for example the English object question in (1). The grammatical object is *who*, but in (1) it occupies the initial position and therefore precedes the subject *the fairies*. But the canonical order of subject and object in a declarative is the reverse: *The fairies are catching the queen*. Much research starting with Chomsky (1957) has argued that *who* in (1) is related to the post-verbal position. We specifically assume an account of (1) involving a movement transformation, though we hasten to acknowledge that approaches involving coindexation, copying, and structure sharing instead of movement also exist and our study was not designed to distinguish between these different theoretical analyses of adult grammar.


¹ARC Centre of Excellence in Cognition and its Disorders, and ZAS (Center for General Linguistics), Germany, ²University of Cyprus and CAT, Cyprus, ³University of Milano-Bicocca, Italy, ⁴University of Belgrade, Serbia, ⁵University of Tallinn, Estonia, ⁶Bar-Ilan University, Israel, ⁷University of Bucharest, Romania, ⁸NOVA University of Lisbon, Portugal, ⁹Vytautas Magnus University of Kaunas, Lithuania, ¹⁰University of Aalborg, Denmark, ¹¹University of Malta, ¹²University of Warsaw, Poland, ¹³University of Groningen, The Netherlands, ¹⁴University of Zagreb, Croatia, ¹⁵University of Salzburg, Austria, ¹⁶University of Oulu, Finland, ¹⁷Warsaw School of Economics, Poland, ¹⁸Catholic University of Louvain, Belgium, ¹⁹University of Athens, Greece, ²⁰ZAS (Center for General Linguistics), Germany

Corresponding author:

Uli Sauerland, ZAS, Schützenstr. 18, 10117 Berlin, Germany
Email: uli@alum.mit.edu

- (1) Who are the fairies catching ___? 

Movement relations are a central interest in theoretical syntax, but also important for many related fields. The study of movement relations in child grammar has provided important evidence for theorists to rely on. Acquisition research has corroborated theoretical syntax with the finding that children find object questions like (1) more difficult to understand than subject questions such as (2) (Seidl, Hollich, and Jusczyk 2003; van der Lely and Battell 2003, and others).¹

- (2) Who ___ is catching the fairies? 

A difference in difficulty between (1) and (2) is predicted by syntactic theories that assume that *who* and the verbal argument position in (1) must be related in some way. Because the relationship between object *who* in (1) and its argument position is longer than that between subject *who* and its argument position, (1) is predicted to be more difficult since the wh-phrase *who* must be held in memory for longer. In this study, we provide further evidence for this prediction of syntactic theory.

We address two further theoretical issues. The first is how the properties of the moved phrase and the subject affect child success at understanding object questions across different languages. To do this, we compare *who* with *which*- phrases like *which princess* relating to both subject and object positions. This comparison also allows us to address the proposal by Adani, van der Lely, Forgiarini, and Guasti (2010); Friedmann, Belletti, and Rizzi (2009); Guasti, Branchini, and Arosio (2012) and others that the difficulty of object movement in children is caused by a difference between adults and children in the syntactic principle *Relativized Minimality* (Rizzi, 1990, 2004, 2013) or related conditions such as the *Minimal Link Condition* (Chomsky, 1995). The Relativized Minimality account predicts that the difficulty of object extraction as in (1) should depend on properties of the moved object phrase as we discuss below. The proposal is controversial (Goodluck, 2010; Varlokosta, Nerantzini, & Papadopoulou, 2015), and some of the original studies claimed to support it worked with very small sample sizes (15 and 22 participants in Avrutin 2000 and Friedmann et al. 2009 respectively), so it is important to have new, stronger evidence to evaluate the proposal.

Secondly, we investigate how sensitive children are to different cues that indicate that the fronted phrase is to be related to the object position. In English, the differences between (1) and (2) are verbal agreement and the position of verb relative to the object. But languages differ in the number of disambiguation cues they use. Consider for comparison the Serbian subject and object question in (3-a) and (3-b). Serbian differs in two ways from English: Case marking on the initial interrogative pronoun (and also on the other noun phrase in the sentence) distinguishes between the subject and object questions, while the word order of the verb and the other argument does not. Some work has already shown that

different disambiguating factors affect children’s performance differently (Adani et al. 2010; Arosio, Yatsushiro, Forgiarini, and Guasti 2011; Guasti, Stavrakaki, and Arosio 2012; Varlokosta, Nerantzini, and Papadopoulou 2015), but no study has yet compared more than two languages.

- (3) SERBIAN
- a. Ko gura princeze?
 who.NOM push.3SG princesse.ACC.PL.FEM
 ‘Who is pushing the princesses?’
- b. Koga guraju princeze?
 who.ACC push.3PL princess.NOM.PL.FEM
 ‘Who are the princesses pushing?’

Our study is also of practical importance for the diagnosis of atypical child language in Europe. Several studies have established that understanding of object questions as well as similar structures (e.g. object relative clauses) is impaired in specific language impairment (SLI, van der Lely and Battell 2003 and others). For this reason, several diagnostic tests for SLI employ structures such as object questions (e.g. the GAPS test, Gardner, Froud, McClelland, and van der Lely 2006; van der Lely, Gardner, McClelland, and Froud 2007) or object relative clauses (e.g. the TROG test, Bishop 2003). But the test designs are primarily based on research on English. At the same time, researchers and practioners have in some cases created adaptations of the tests for other languages. For many languages with fewer speakers than English, such an adaptation of the fruits of the research on English can be an effective strategy to meet a demand for language diagnostics. But the strategy carries a risk if we don’t know how the individual properties of a specific language affect the acquisition process and interact with language impairment. It could be for example that the presence of case marking in Serbian changes how sensitive question understanding tests are to SLI. In this study, we investigate the influence of grammatical variation on the understanding of questions in typical five year olds across Europe.

Methods

Design

The two factors investigated within each language were whether the wh-phrase is simple (in English, *who*) or complex (in English, e.g. *which princess*) and whether it is the subject or object argument of the verb. This leads to a 2×2 design and we created 6 items for each of the four conditions. Sample items in English are shown in (4). All English items and sample items for all 18 languages are presented in the two appendices.

- (4) a. Who — is pushing the princesses? (subject, simple)
 □

- b. Which lady — is pushing the queens? (subject, complex)
└──────────┘
- c. Who are the dancers pushing —? (object, simple)
↑──────────────────┘
- d. Which princess are the ladies pushing —? (object, complex)
↑──────────────────┘

The lexical material used was selected to make it possible to translate the items into the 18 languages under study without introducing confounds for the cross-linguistic comparison. In particular, we avoided items that would be ambiguous in one of the target languages in a relevant way. Such an ambiguous translations would have arisen, for example, if we had used items with a singular definite as illustrated by German (5). (5) is globally ambiguous between the subject question and the object question interpretation that are distinguished in English by the position of the verb.

- (5) Welche Frau schiebt die Prinzessin?
 which woman pushes the princess
 ‘Which lady is pushing the princess?’ / ‘Which lady is the princess pushing?’

With the plural definite, verbal agreement disambiguates between the two interpretations as shown in (6). When the verb has singular agreement only the subject question interpretation is possible, and with plural agreement only the object question interpretation. All items with two lexical nouns like (4-b) and (4-d) contained a number contrast. Namely, the fronted question phrase was singular, and the other argument of the verb was marked plural. Because of the number contrast, verbal agreement for number could serve as a cue as to whether a subject or object interpretation was correct at least in these items (i.e. those of the complex condition).

- (6) a. Welche Frau schiebt die Prinzessinen?
 which woman push.SG the princesses
 ‘Which lady is pushing the princesses?’
 b. Welche Frau schieben die Prinzessinen?
 which woman push.PL the princesses
 ‘Which lady are the princesses pushing?’

Furthermore, we ensured that the items were uniform in each language with respect to morphological marking for case and gender, and verb complexity. We chose lexical nouns that are morphologically marked as the same gender in all languages; namely, feminine in languages with masculine, feminine, and neuter gender, and as common gender in languages with only common and neuter gender. The six nouns we used are *princess*, *queen*, *granny*, *lady*, *fairy*, and *dancer* in English. Furthermore six verbal actions were chosen so that two criteria would be satisfied for all 18 languages involved. The first criterion was that the actions could be described by a monomorphemic verb root in all 18 languages. The

second criterion was that all six verbs exhibit uniform case properties within each of the languages. Since languages such as Finnish and Lithuanian exhibit a rich case system, we made sure that in all examples the case marking of the object arguments was identical. The six verbs used are *push*, *pull*, *wash*, *catch*, *feed*, and *scratch* in English. In some languages, the verb for *tickle* was used instead of the one for *scratch*.

The items were created simultaneously in the 18 languages in this study: Estonian, Finnish, Maltese, Hebrew, Greek (both Standard and Cypriot), Lithuanian, Croatian, Serbian, Polish, English, German, Dutch, Danish, French, Italian, European Portuguese, and Romanian.² For each language, we used a general, natural way of asking constituent questions, but within each language the same pattern was used consistently for all items of the same type. The second appendix presents a set of four sample items for each language. In addition to the different cues (word order, case, agreement), variation between languages exists in at least three other respects: 1) whether a synthetic or analytic verb form is used, as illustrated by the contrast between the analytic English progressive forms like *is pushing* in (4) and the synthetic Serbian present tense form like *gura* ('push') in (3); 2) whether a cleft structure was used for the question in the European Portuguese example (7); and 3) whether the object questions used a clitic pronoun doubling an argument as in French (8).

- (7) Quem é que está a puxar as senhoras?
 who is-3SG that is-3SG at pulling the-PL ladies-PL?
 lit. 'Who is (it) that is pulling the ladies?' ('Who is pulling the ladies?')
- (8) Qui les grands-mères grattent-elles?
 who the grand-mother.PL scratch-they
 'Who do the grannies (they) scratch?'

Children's understanding of the questions was tested in a picture choice task. The same pictures were used for all 19 test sites. The samples in figure 1 illustrate this aspect of the task. In addition to the target picture three alternative pictures were shown as choices corresponding to three different types of misunderstandings. Specifically, these were: 1) A number error picture corresponding to the question where the number of the non-questioned argument is singular rather than plural as in (9-b). Note that, in object questions children committing a number error, also ignore number agreement on the verb which Johnson, de Villiers, and Seymour (2005) show that five years olds are sensitive to in other cases. 2) A reversal error picture corresponding to the question where the grammatical roles of subject and object are switched as in (9-c). Because the subject and object phrases also differ in number, the reversal error entails also an error on the number agreement of the verb, but not the number marking on the nominals. 3) A semantic verb error corresponding to a question where verb is different from the verb in the actual question as in (9-d).

- (9) a. actual question: Who are the queens washing?
 b. number error: Who *is* the *queen* washing?

- c. reversal error: Who *is washing the queens*?
- d. semantic verb error: Who are the queens *chasing*?



15: Who are the queens washing?

19: Which princess are the dancers pushing?

Figure 1. Sample picture choice displays for items 15 on the left and 19 on the right. For item 15, the four possible responses correspond to positions as follows: 'CORRECT' is top left, 'REVERSAL ERROR' is bottom left, 'NUMBER ERROR' is bottom right, and 'SEMANTIC VERB ERROR' is top right. For item 19, the correspondence is: 'CORRECT' is top right, 'REVERSAL ERROR' is bottom left, 'NUMBER ERROR' is top left, and 'SEMANTIC VERB ERROR' is bottom right.

At each test site, all items were recorded by a native speaker of the local variety of the target language. Two versions of the experiment were created as a Microsoft Powerpoint presentation. The two versions only differed with respect to the order of the experimental items, and we refer to them as order A and B in following. The experiment included three initial slides as part of the instructions to the child. The first slide introduced all the six types of characters involved in the slides. The second and third slides were practice items. The experiment was presented to the participants on a laptop computer using the internal speakers of the laptop. Participants were asked to indicate their answer to the question by a hand gesture, which was recorded by the experimenter. The results were then coded at each site in an anonymized format and collected centrally for the analysis.

Participants

Both children and adults participated in the experiment. The 392 child participants were tested in 19 different countries with 18 different languages. The age range was 4;10 to 6;0. The number of child participants and language, age, and gender information at the individual test sites is shown in table 1.

In the following, we frequently use the two letter ISO country codes shown in the first column of table 1. The number of child participants varied between 17 (Lithuania and Israel) and 25 (Austria). In each country at least 7 children

ISO	country	language	subfamily	n	female	m(age)	SD(age)	%order A
EE	Estonia	Estonian	Finnic	20	10	5;5.5	3.5	50
FI	Finland	Finnish	Finnic	20	12	5;4.9	3.2	50
MT	Malta	Maltese	Semitic	20	11	5;7.0	3.1	50
IL	Israel	Hebrew	Semitic	18	11	5;7.2	3.0	50
CY	Cyprus	Cypriot Greek	Greek	20	10	5;3	2.7	50
EL	Greece	Greek	Greek	21	11	5;5.8	4.0	52.4
LT	Lithuania	Lithuanian	Baltic	17	7	5;7.9	2.2	58.8
HR	Croatia	Croatian	Slavic	21	11	5;5.8	3.2	47.6
RS	Serbia	Serbian	Slavic	20	12	5;5.0	3.3	50
PL	Poland	Polish	Slavic	22	12	5;8.2	3.3	50
UK	Great Britain	English	Germanic	20	10	5;4.0	2.6	50
DE	Germany	German	Germanic	20	10	5;7.7	3.3	50
AT	Austria	German	Germanic	25	10	5;6.2	3.8	52
NL	Netherlands	Dutch	Germanic	20	10	5;7.3	3.4	45
DK	Denmark	Danish	Germanic	21	8	5;4.7	2.7	47.6
BE	Belgium	French	Romance	24	14	5;3.8	3.3	50
IT	Italy	Italian	Romance	20	10	5;2.4	3.5	50
PT	Portugal	E. Portuguese	Romance	20	13	5;5.2	3.0	50
RO	Romania	Romanian	Romance	23	10	5;5.8	2.5	47.8
total	19	18	7	392	202	5;5.6	3.5	50

Table 1. Overview of country and language distribution of 392 child participants.

of either gender participated. The mean age of children differed by almost 6 months between the youngest (Italy) and the oldest (Poland) group. The age differences are in part explained by the variation in school entry age across the European countries and whether our collaborating educational institution was a pre-school or a primary school. In the statistical analysis, we therefore include age as a factor of the analysis to differentiate the effects of age and language. Finally, order A and order B of the list of test sentences are close to balanced in each country with a 58.8 : 41.2 balance in Lithuania the furthest away from the ideal 50:50 balance. In addition to the children, 59 adults (38 females, 21 males) participated in the study in 6 countries. The number of adult participants by country was 10 Estonia, 11 Croatia, 10 Lithuania, 10 Malta, 5 Netherlands, and 13 Poland. Unfortunately it wasn't possible to obtain adult data from all 18 languages in this study, but we still found the adult data useful to exclude five problematic items.

Linguistic Factors

To capture the effects of cross-linguistic variation, we coded specific linguistic properties that varied across the languages in our sample. The seven linguistic factors that we included in the cross-linguistic analysis are agreement (agr), word-order (wo), distinctive case morphology (case), one wh-marker (1wh), cleft

fam. lang	Finnic		Sem.		Greek		Bl.	Slavic			Germanic			Romance				
	EE	FI	MT	IL	CY	EL	LT	HR	RS	PL	UK	DE,AT	NL	DK	BE	IT	PT	RO
case	●	●	●	●	●	●	●	●	●	●	○	●	○	○	○	○	○	●
wo	●	●	○	●	○	○	○	○	○	○	●	○	○	●	●	○	●	○
agr	●	●	●	●	●	●	●	●	●	●	●	●	●	○	●	●	●	●
lwh	○	●	○	●	●	●	●	●	●	●	○	○	○	○	●	○	●	○
clf	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●	○	●	○
syV	○	○	●	○	○	○	○	○	○	○	●	○	○	●	○	○	●	○
CID	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	●
pro	●	●	●	●	●	●	●	●	●	●	○	○	○	○	○	●	●	●

Table 2. Table of linguistic factors considered in the analysis.

structure (clf), synthetic verb form (syV), clitic doubling (CID), and subject pro drop (pro). The values of the eight factors are shown in the table 2, and we explain in more detail in the following how we arrived at these value. All factors are binary at the item level. The mark ● in table 2 indicates that the factor is true for all items of the language, the mark ○ indicates the opposite. The mark ● indicates that the items of the respective language are not uniform with respect to the factor, e.g. simple wh-phrases in some cases differ from complex wh-phrases. The items used in Germany and Austria were identical in the written form, so there is only one column for the two countries. The voice recordings for the two countries were different though, to match the local pronunciation.

Consider in more detail the eight factors represented in the table in the order given. Agreement (agr) represents whether number agreement of the verb is distinct for the subject questions and object questions, and thus could serve as cue to understand the question correctly. In two languages (Italian and Dutch) agreement is the only cue. Dutch is unusual: The Dutch equivalent of English *‘Who are singing?’ is grammatical if the expected answer is a plurality. As a consequence, the Dutch translation of the object question given in (10) outside of a specific situation is ambiguous between a subject and an object question interpretation. In the experimental situations, however, it was always disambiguated to a subject question interpretation by the fact that none of the four pictures matched the object question interpretation. But since the agreement marking itself doesn’t disambiguate (10) syntactically, the factor agreement was false for Dutch *who* questions. But Dutch complex singular wh-phrases such as *welke prinses* (‘which princess’) are disambiguated by number agreement of the verb. This is represented by ● in table 2.

- (10) Wie trekken de prinses-sen?
 who pull.3PL the princess-PL
 ‘Who is pulling the princesses?’ / ‘Who are the princesses pulling?’

In Danish, we used question structures where verbal agreement doesn't distinguish between subject and object questions.

Word order (wo) indicates whether the order of the main verb and the definite noun phrase distinguishes between the subject and object questions. In our sample verb-medial (SVO) languages like English (see example (4)) have the entry ● here, while the entry for verb-final (SOV) languages like German (see example (5)) is ○.³ Case (case) captures whether there is overt case marking that distinguishes between subject and object questions. The table entry for English here is ○ because the form of the wh is always *who* or *which*.⁴ Serbian as shown in (3) is a language with case morphology. The entry for German is ● because the simple wh-forms *wer* vs *wen* in German show case morphology, but the complex forms in the feminine gender don't.⁵

The factor 1wh indicates whether the simple and complex wh-phrases use the same morpheme or two different ones. The entry for English is ○ because the morphemes *who* and *which* are distinct. But in Finnish, the simple wh-word *kuka* ('who') is exactly the same as the form used for complex wh-phrases such as *kuka keiju* ('which fairy'). In Greek and Cypriot Greek the two forms are not exactly identical as used in the examples, but this is because the simple form *pjos* ('who'/'which.M') corresponds to the masculine singular agreement while the feminine form *pja* ('which.F') is used as part of the complex phrases. The factor clf (cleft) indicates whether the question used had the structure of a cleft (see the Portuguese (7) above). In French, the subject questions used a cleft structure, but not the object questions. The factor syV (synthetic verb form) shows whether the verb form consisted a single (inflected) verb or involved auxiliaries. In English, the use of the progressive forms *is/are pushing* is represented by the ● entry in the table. The factor CID (clitic doubling) indicates whether any argument phrase in the question was doubled by a pronoun (see (8) above). In Romanian, clitic doubling is optional with simple wh-phrases and obligatory with complex (D-linked) wh-phrases. We used clitic doubling only with the latter, hence the table entry ●. Finally the factor pro indicates whether the language in question permits the omission of subject pronouns (i.e. the pro drop parameter of [Chomsky 1981](#)). For example, in English omission of *she* from '*She is sneezing.*' is ungrammatical, and it has the entry ○. None of our experimental items contained a subject pronoun, so a possible effect of factor pro would have to be a language level effect. Glossed sample items for all 18 languages are provided in the second appendix.

Statistical Analysis

Statistical analyses were performed using R ([R Core Team, 2012](#)) and mixed logit models ([Bates et al., 2015](#)) for each response type. We used a stepwise forward inclusion procedure, starting with the null model with only the random factors Subject and Item. We then added one potential predictor at a time and comparing the model with the predictor against the one without it, using a χ^2 test ([Jaeger 2008](#)). We first tested whether presentation position or the verb used would significantly improve model fit. Then the fixed factors age

(in months), argument type, wh-type, case, agr, and wo were considered in this order in all analyses. For the analysis of correct responses, we furthermore introduced the other five linguistic factors from table 2 in the order 1wh, clf, syV, CID, and pro. In all the models, the reference category for the predictor Argument type is object questions, for Wh-type is Which-phrase, and for the linguistic factors of table 2 the value marked \circ in the table. Both first-level effects and interactions between the fixed-effect factors were examined. Thus, we established which factors contributed significantly to the model's fit. Based on z values (Wald statistics), we obtained an estimation of the statistical significance of each predictor in the model.

Mixed logit models do not provide an intuitive measure of effect sizes but the coefficients we report are the log of the odds-ratio due to the predictor (Hosmer Jr., Lemeshow, and Sturdivant 2013, ch. 3, and others). For example, the coefficient 0.544 for predictor Wh type in table 3 shows that the odds of a correct response are $\exp(0.544) = 1.723$ times higher for simple than for complex wh-phrases. For reader convenience, we report in table 3 also individual percent correct ratios (ind %cr) for the presence/absence of the significant predictors of correctness as an intuitive measure. These values have no direct interpretation in our multifactorial model, but provide an intuitive measure of the effects of the predictors. Take the predictor Wh type in table 3 as an example: the individual probabilities correspond to odds of 644 : 366 = 1.760 and 513 : 487 = 1.053 and the ratio of these two odds, 1.671. This value is similar to the exponent of the coefficient of the logit model for Wh type, 1.723, but because the logit model takes into account other factors the two values aren't identical.

Results

We obtained 24 responses per subject. Items 4, 10, 13, 16, and 22 were excluded from the further analysis for the following reasons: Items 4, 10, 16, and 22 used the verb *catch* and required a distinction between a picture depicting *pull* and one depicting *catch*. They showed adult semantic verb error rates up to 35.6%. Item 13 showed an even higher rate (40.7%) of semantic verb errors in the adult results. After these exclusions, the number of correct responses by adults was 96.8% (MT 94.9%, LT 100%, HR 95%, PL 98.5%, NL 89.8%).

In the following, we analyze the data from 392 children on the 19 remaining items, which amounts to 7446 responses because two responses are missing. Figures 2 and 3 show the frequency of the four possible responses across the 19 sites. Figure 2 compares the subject questions with the object questions. There are two languages where the object questions elicited more correct responses than the subject questions, namely, Hebrew and Polish. At the other 17 sites, subject questions elicited more correct responses with the greatest difference in Dutch. The number of reversal errors also shows an interesting pattern since it is higher for object questions than for subject questions at all 19 sites.

Figure 3 displays the same data as figure 2, but as a comparison between simple and complex questions. The number of correct responses for simple questions is higher than for complex questions across all 19 countries. But the

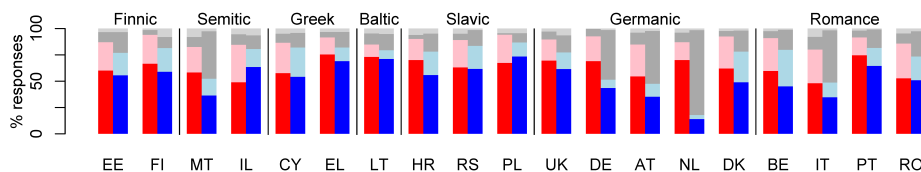


Figure 2. Comparison of subject (red) vs object (blue) questions across 19 countries. Correspondences: Dark red/blue – correct, light red/blue – number error, dark grey – reversal error, light grey – semantic verb error

effect of the factor simple vs complex seems to vary less across languages than that of subject vs object. For more detailed findings, we proceed to the statistical

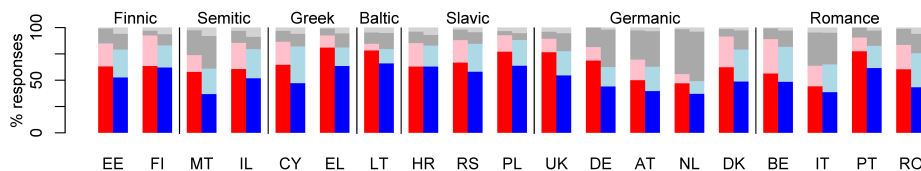


Figure 3. Comparison of simple (red) vs which (blue) questions for 19 countries. Correspondences: as in figure 2

In the first model we computed, Correct response was the dependent variable. Following the stepwise inclusion procedure described above, we first verified that neither the presentation position of an item nor the verb used were significant factors. Of the other fixed factors, the following factors contributed significant information and were included in this order where we show the significance of improvement in model fit for each step: Age [$\chi^2(1) = 14.08$, $p < .001$], Argument type [$\chi^2(1) = 6.18$, $p = .01$], wh-type [$\chi^2(1) = 11.55$, $p < .001$], and case [$\chi^2(1) = 10.11$, $p < .01$]. An interaction Argument type by case was also significant [$\chi^2(1) = 43.07$, $p < .001$]. Part a of table 3 indicates that the probability of correct responses increases with age; it also increases from object to subject questions (positive coefficient, 0.93), from complex to simple questions, from Questions lacking morphological Case to questions with morphological Case. To unpack the interaction (Sentence Type by Case), we ran separate analyses for subject and object questions. Case provided significant information in object questions [$\chi^2(1) = 36.58$, $p < .001$]: Object questions with Case were better understood than object questions without Case. This is illustrated in Table 3, Part b. The probability of correct responses increases from object questions without Case to object questions with Case. Word order (wo) was also found to significantly contribute to model fit [$\chi^2(1) = 10.83$, $p < .001$]. Finally, two further linguistic factors contributed significantly: 1wh [$\chi^2(1) = 31.11$, $p < .001$] and synthV [$\chi^2(1) = 10.51$, $p < .002$]. Table 3 shows that the probability of a correct response increases if words for *who* and *which*

are identical and if the questions used a synthetic verb form. In the model we had thus arrived at, the factor *wo* didn't contribute a significant effect anymore, and we verified that the fit of a model not including *wo* as a predictor didn't fit significantly worse than the one that included *wo* [$\chi^2(1) = 0.64, p = .42$]. Hence the final model we report below excludes *wo*.

Table 3. Summary of the fixed effects in the mixed logit models for children for Correct responses

Predictor	Coeff.	SE	Wald Z	p	ind %cr
Part a: (N = 7446, children = 392, item = 19, loglikelihood = -4692.2)					
(Intercept)	-4.06226	0.79870	-5.086	<.001	
Age	0.04776	0.01199	3.984	<.001	[55.0%:60.1%] ⁶
Argument type	0.94123	0.16463	5.717	<.001	62.9%:51.5%
Wh type	0.54416	0.10306	5.280	<.001	64.4%:51.3%
Case	0.57532	0.15127	3.803	<.001	61.0%:41.7%
1wh	0.61446	0.08875	6.924	<.001	63.3%:51.3%
syV	0.42762	0.11075	3.861	<.001	59.3%:57.0%
Arg. type by Case	-0.69618	0.10554	-6.596	<.001	[see text]
Note. Random effects for participants and items had SD of 0.64 and 0.31, respectively					
Part b: Object questions (N = 3527, children = 392; item = 9, loglikelihood = -2275.1)					
(Intercept)	-0.35	0.16	-2.12	.03	
Case	0.65	0.11	5.80	<.001	57.6%:41.7%

As discussed above, column 'ind %cr' of table 3 shows the individual ratio of correct response percentages as an intuitive measure of effect size. For the factors argument type and case that also interact with each other, the picture is more complicated. For the significant interaction between argument type subject and case the corresponding probabilities are a 2 by 2 matrix with the following values: argument type subject with case morphology, 63.4%, without case 62.2%, object with case: 57.6%, and object without case: 41.7%. Table 4 shows the individual %correct ratios for the factors that weren't significant predictors of correct responses.

Table 4. Individual percent correct for non-significant factors

Factor	wo	CID ⁷	pro	cleft
ind %cr	57.0%:58.1%	44.2%:52.2%	59.9%:52.6%	69.5%:56.9%

Notable about table 4 is that the last three non-significant factors exhibit ratios more extreme than some of the significant predictors. But note that all three also correlate strongly with some of the significant predictors as can be seen in table 2. Specifically, *clf* in French and *CID* in both French and Romanian correlate with Argument type. Factor *pro* matches factor Case for 15.5 out of

18 languages,⁸ and Case exhibits an individual ratio more skewed than pro. To verify that including Case rather than pro as a factor in the model is justified, we first compared the significance of the two at the step of the inclusion procedure where Case was added, but in both cases the significance was so great that the software we used showed no difference. We then computed a model including factor pro in addition to the predictors of the model described in table 3. Since we found pro was not a significant predictor, while Case remained a significant predictor, we decided to report the model including Case. Factor pro also matches factor 1wh for 13 out of 18 languages, but 1wh exhibits an individual ratio more skewed than pro too. To further confirm that inclusion of 1wh over pro yields a better model, we compared the two at the step of the inclusion procedure at which we tested 1wh. Including pro instead at this point also would improve model fit, but the significance of the improvement was much lower ($\chi^2(1) = 6.37, p = .012$ for pro vs. $\chi^2(1) = 31.1, p < .001$ for 1wh).

In the error analysis, we didn't consider the linguistic factors other than Case, Agreement (agr), and Word order (wo) because we had no specific hypothesis for these. Children made three different types of errors: Reverse, Number, and Semantic errors. In the second model with Reverse response as the dependent variable in table 5, the following factors contributed significant information and were included: Argument type [$\chi^2(1) = 34.96, p < .001$], Wh-type [$\chi^2(1) = 5.09, p < .05$], Case [$\chi^2(1) = 36.30, p < .001$], and wo [$\chi^2(1) = 55.22, p < .001$]. The interactions Argument type by Case [$\chi^2(1) = 29.76, p < .001$], Argument type by wo [$\chi^2(1) = 47.83, p < .001$], Wh-type by wo [$\chi^2(1) = 8.23, p < .01$], and Case by wo [$\chi^2(1) = 22.20, p < .001$] were significant. To unpack the first two interactions we ran separate analyses for Subject and Object questions and found that Case and wo mattered only for object questions. The probability of making a reverse error decreases from object questions without case to object questions with case, and from object questions with -wo to object questions with +wo, as evident in table 5 part b. Then, we ran separate analysis for wo. Predictor argument type was significant for both +wo and -wo languages shown in parts c and d of table 5. Subject questions were less likely to receive a reverse response in both language types, but the effect was weaker in +wo languages. Wh-type was only significant for -wo languages, where Simple questions were less likely to receive reverse responses. Case, however, only was significant for +wo languages, where the combined absence of wo and Case saw more reverse responses.

Table 5. Summary of the fixed effects in the mixed logit models for children for Reverse responses

Predictor	Coeff.	SE	Wald Z	p
Part a: (N = 7446, children = 392, item = 19, loglikelihood = -2798.4)				
(Intercept)	0.33	0.16	2.02	.04
Argument type	-2.66	0.19	-13.7	<.001
Wh-type	-0.06	0.16	-0.34	.73
Case	-1.65	0.14	-12.2	<.001
wo	-1.54	0.17	-9.15	<.001

Table 5. continued from previous page

Predictor	Coeff.	SE	Wald Z	p
Argument type by Case	1.07	0.16	6.85	<.001
Argument type by wo	1.18	0.16	6.87	<.001
Wh-type by wo	-0.52	0.16	-3.35	<.001
wo by Case	1.09	0.22	4.94	<.001
Note. Random effects for participants and items had SD of 0.73 and 0.29, respectively				
Part b: Object questions (N = 3527, children = 392, items = 9, loglikelihood = -1750.9) for Reverse responses.				
(Intercept)	-0.28	0.14	1.99	.047
Case	-1.75	0.16	-10.45	<.001
wo	-1.71	0.22	-7.93	<.001
Note. Random effects for participants and items had SD of 1.02 and 0.0001, respectively.				
Part c: Questions with +wo (N = 2717, children = 249, items = 19, loglikelihood = -928.2) for Reverse responses.				
(Intercept)	-1.44	0.17	-8.41	<.001
Argument type	-1.04	0.19	-5.56	<.001
Wh-type	-0.54	0.19	-2.85	<.01
Case	-1.7	0.17	-1.02	.31
Note. Random effects for participants and items had SD of 0.64 and 0.29, respectively.				
Part d: Questions with -wo (N = 4729, children = 143, items = 19, loglikelihood = -1895.8) for Reverse responses.				
(Intercept)	0.01	0.17	0.54	.59
Argument type	-2.03	0.17	-11.7	<.001
Wh-type	-0.05	0.17	-0.31	.76
Case	-1.3	0.13	-10.24	<.001
Note. Random effects for participants and items had SD of 0.77 and 0.31, respectively.				

In the third model with Number response as the dependent variable in table 6, the following factors contributed significant information and were included: Age [$\chi^2(1) = 14.29, p < .001$], Argument type [$\chi^2(1) = 7.58, p < .01$], Wh type [$\chi^2(1) = 5.72, p < .05$], and wo [$\chi^2(1) = 8.77, p < .01$]. The interactions Argument type by wo [$\chi^2(1) = 7.98, p < .01$] and Wh type by wo [$\chi^2(1) = 3.99, p < .05$] were also significant. To unpack them, we ran separate analyses for +wo and -wo languages. The two predictors Argument type and Wh type were only significant in -wo languages. Specifically, subject questions received more number error responses than object questions, and simple wh-phrases received more number error responses than complex wh-phrases in -wo languages as evident in part b of table 6.

Table 6. Summary of the fixed effects in the mixed logit models for children for Number Error responses (N= 7446, children = 392, items = 19, loglikelihood = -3776.5)

Predictor	Coeff.	SE	Wald Z	p
Part a				
(Intercept)	1.23	0.81	1.5	.13
Age	-0.04	0.01	-3.61	<.001
Argument type	0.61	0.13	4.5	<.001
Wh type	-0.32	0.12	-2.57	<.05
wo	0.44	0.11	4.0	<.001
Argument type by wo	-0.33	0.11	-2.85	<.001
Wh type by wo	0.24	0.12	2.02	<.05

Note. Random effects for participants and items had SD of 0.61 and 0.24, respectively

Part b: -wo languages (N = 4729, children = 249, items = 19, loglikelihood = -2265.8) for Number error responses.

(Intercept)	2.16	1.05	2.05	<.05
Argument type	0.62	0.16	3.82	<.001
Wh type	-0.43	0.16	-2.65	<.01

Note. Random effects for participants and items had SD of 0.67 and 0.31, respectively.

In the fourth model, with verb semantic error response as the dependent variable, only Case was a significant predictor [$\chi^2(1) = 12.06, p < .001$]. As table 7 shows, more verb semantic errors occurred in +case questions than -case questions.

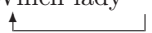
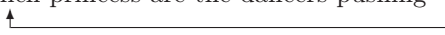
Table 7. Summary of the fixed effects in the mixed logit model for children for Verb Semantic Error responses (N= 7446, children = 392, items = 19, loglikelihood = -1051.1)

Predictor	Coeff.	SE	Wald Z	p
(Intercept)	-4.37	0.33	-13.2	<.001
Case	0.51	0.14	3.57	<.001

Discussion

Our results from 18 different languages show that there is substantial variation, but also some uniformity across languages in the understanding of questions by 5 year olds. For all 18 languages in the study, we found that object questions elicit more reversal errors than subject questions – i.e. errors where children mix up the grammatical roles of subject and object. This result also corroborates many earlier findings, starting with Thornton (1990), that object questions are more difficult for children than subject questions. Moreover, in our data the percentage of correct responses exhibits a similar pattern, but not quite as universal. Only Hebrew and Polish children gave more correct responses to object questions than to subject questions. We think that the effect of argument type on correct responses may be a little weaker than on reverse responses because in our design the number error response could only be given by

children who assigned the argument roles correctly. Since the statistical analysis indicates that argument type is a significant predictor of correct response, we think the reverse result from Hebrew and Polish is due to random variation, and would not expect it to hold up in further testing. Our result of Argument type is expected because of the structural properties of the languages in our sample which was determined by funding constraints rather than scientific considerations. Specifically, all languages in our sample had two important structural characteristics: 1) the most frequent word-order in declaratives places the subject (S) in front of the object (O) (either SVO or SOV), and 2) in questions the *wh*-phrase is fronted to the left periphery of its clause. Therefore, subject questions corresponded in all languages of our sample with the most frequent order of subject and object, while the order of object questions was the reverse of that. There are therefore several factors that could make object questions more difficult for children than subject questions. For example, the deviation from the most frequent word order could be the reason. But the reason might also be the length of the dependency between the *wh*-phrase and its trace position (e.g. 0 words between *wh*-phrase and trace in English (11-a), but 4 in (11-b)) is longer for object questions. Or the reason could be that specifically the presence of the subject in the movement path of the object interferes with syntactic movement of the object for some children.

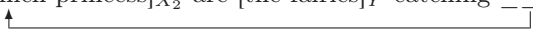
- (11) a. Who / Which lady — is pushing the princesses?

- b. Who / Which princess are the dancers pushing —?


The second finding true of all 18 languages is that more correct responses are given to questions with a simple *who*-phrase than to questions with a complex *which*-phrase (see figure 3). This finding corroborates earlier results by Avrutin (2000) and Goodluck (2010). Avrutin and Goodluck discuss this finding in terms of the semantic property of D-linking (Pesetsky, 1987). Our crosslinguistic data indicate that another factor is relevant too. Namely, we found that the factor 1wh, indicating morphological identity of the words for *who* and *which* in a language, predicts higher rates of correct responses. This finding suggests that children speaking a language like English are less familiar with the word *which* than the word *who*, while in languages like Finnish where *kuka* is the word for both *who* and *which* this difficulty doesn't arise. But, the semantic factor D-linking seems to be also relevant since we consistently find better performance on *who*-questions even in the 10 languages that are like Finnish in this respect. In sum, the two universal patterns are for the most part expected, the effect of the identity of *who* and *which* is novel.

We now come to predictors of the variation. We found three language-specific factors that predicted a greater probability of correct response: the presence of over Case-marking on the moved phrase, whether there is only a single morpheme for both *who* and *which* in the language mentioned in the previous paragraph already, and whether the verb form used is synthetic.

Furthermore we showed that some other plausible linguistic factors were not significant predictors: whether a cleft structure was used or not, whether a pronoun doubling the *wh*-phrase was used or not, and whether the language allowed subject pronoun to be omitted. For clefts and pronoun doubling, our data set only included two languages each that had the phenomenon, so these may warrant further testing. Of the significant predictors, *syV* wasn't considered in any previous studies. The four languages with the auxiliary show little commonality though: In English and Portuguese, the auxiliary exhibits number agreement, but in Maltese and Danish it doesn't, and only in Maltese is the main verb marked for agreement. We think nevertheless that English children may benefit from the early presence of the number agreement cue in examples starting with *who is/are . . .* when determining the question interpretation, and Portuguese may be similar. Note that these two languages show markedly better comprehension than all other languages that lack relevant case morphology.

A closer look at the other factors of variation we found is particularly interesting from the perspective of the relativized minimality proposal of [Grillo \(2009\)](#); [Rizzi \(2004, 2013\)](#) and [Friedmann et al. \(2009\)](#). Relativized minimality proposes that the interpretability of movement of a phrase depends on relationships between the syntactic features of any intervening phrase (i.e. c-commanded by the moved phrase and c-commanding the trace) and those of the moving phrase. Generally match of some or all features of the two relevant phrases is assumed to cause interference. To apply this to our data, consider the configuration in (12) where the feature set of the object *wh* phrase is X (X_1 for *who*, X_2 for *which*) and that of the subject is Y . Relativized minimality predicts that the interpretability of the two kinds of object questions depends on the relation of X and Y .

(12) [Who] $_{X_1}$ / [Which princess] $_{X_2}$ are [the fairies] $_Y$ catching ___?


We can use our data to deduce for which specific features there is an effect of relativized minimality. This test relies on the fact that the subject questions we tested never violate Relativized Minimality. Therefore any difference we found between the simple and complex subject questions must be the effect of other properties; e.g. the difficulty of accessing *which* in the lexicon and the D-linked status of *which* that we mentioned above. Any effect of Relativized Minimality therefore must be visible as an interaction with argument type: for example, a difference between the two object questions, but no corresponding difference between the subject questions. The features have been implicated in relativized minimality in other studies are NP ([Friedmann et al., 2009](#)), case ([Arosio, Guasti, & Stucchi, 2011](#); [Arosio, Yatsushiro, et al., 2011](#); [Guasti, Stavrakaki, & Arosio, 2012](#); [Varlokosta, Nerantzini, & Papadopoulou, 2015](#)), number ([Adani et al., 2010](#)), and gender ([Adani et al., 2010](#)). In our study, gender was not tested,⁹ but the features NP, case, and (indirectly) number were relevant. For case, we found an interaction with argument type in the direction predicted by relativized minimality – children are likelier to comprehend object questions if the overt case marking is present. For NP, however, we didn't find the interaction predicted

by relativized minimality. The parameter +NP corresponds in our data to the Wh type = Complex for the *which*-questions, and -NP corresponds to the *who* questions. The presence of an interaction for Case and the absence of one for NP is also visually apparent in figure 4: only for case is the difference between the +Case and -Case probabilities greater for objects than for subjects.

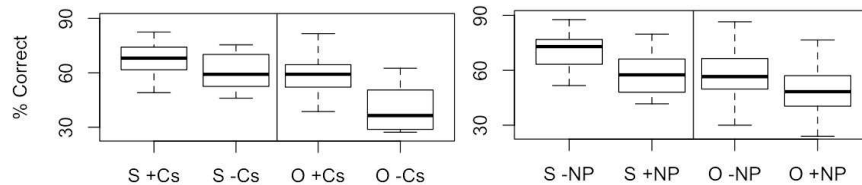


Figure 4. Boxplots of mean % correct per language showing the Case–Argument type (Cs: Case, S: Subject, O: Object) interaction on the left and the absence of a NP–Argument type interaction on the right.

While the effect of case is consistent with the earlier findings mentioned above, the absence of an effect for NP differs from findings reported in the literature with the exception of Goodluck (2010). Since our study with 392 children is more powerful than previous studies (Friedmann et al. 2009 report data from 22 children in their subexperiment 5), the effect found in previous studies should have also been detected in ours. We can think of three possible reasons for the discrepant results: 1) differences in the design especially with respect to number, 2) the effect is restricted to specific populations and languages, and 3) the result reported by Friedmann et al. (2009) is a false positive. Consider first a relevant design difference: As one reviewer points out, our experimental items all involved a number contrast between the singular wh-phrase and a plural in the other argument position. Adani et al. (2010) have argued that a number difference can ameliorate a relativized minimality effect, so the number difference may override any effect of the NP feature in our data. A remaining problem with this proposal, though, is that our data shows that the effect of case cannot be overridden by a number difference in the same way. This isn't predicted by the theory assumed by Friedmann et al. (2009), which is stated in terms of subset/superset relations of feature sets, since the difference in number features should block any subset relationship. To accommodate our data, we could instead assume that the number feature subcategorizes the NP feature as represented by $\pm\text{NP}_{\pm\text{PL}}$ and that number of features in the intersection of the two feature sets correlates with their interpretability. While this isn't implausible, other explanations of the discrepancy should also be considered at this point. Secondly, it might be that the NP effect is restricted to some languages or age groups – Friedmann et al. (2009) tested only 4-year olds speaking Hebrew. We therefore tested the data from all 19 sites individually for an interaction using a logistic regression with the three factors Wh type, Argument type, and their interaction. But, we found that the interaction was significant at no site.¹⁰ Hence, our data shows it to be unlikely that only some

languages are subject to intervention by the feature NP in object questions. The difference in age, however, may be the cause of the difference in findings. Finally, a third possibility for the discrepancy between our result and Friedmann et al.'s (2009) may be that their result is a false positive, as Goodluck (2010) suggests. Friedmann et al. (2009) tested only 22 children and furthermore only report pair-wise *t*-tests of the number of correct responses per subject in the statistical analysis of their data. Since *t*-tests assume normal distributions, their use may lead to false positives if applied to categorical and count data (Jaeger 2008 and others). When we compute a non-parametric statistic, the χ^2 -test, for the counts of 'No. of participants of 22 who performed above chance' of Friedmann et al.'s (2009) table 6, our analysis shows a substantial risk that result may indeed be a false positive ($\chi^2(1) = 2.2, p = .14$).¹¹ At this point, our data cannot decide between the three explanations of the discrepancy, but we hope future research will resolve the question.

In the analysis of reverse responses in 5, we also found effects of Argument type, Wh type, case, and the interaction of Argument type by case, each in the opposite direction compared to the analysis of correct responses. This is expected because of the complementarity between the two types of response. But in addition, word order (wo) was a significant predictor of reverse responses, individually and also in pairwise interaction with Wh type, Argument type, and case.¹² As a single factor disambiguation by word order led to fewer reverse responses which is expected since children are sensitive to word order from a young age on (Seidl et al., 2003). The reason that we don't find a complementary relationship in the case of Correct responses seems to be that children in +wo languages give more Number errors (see table 6 above). The three interactions involving word order are more interesting. First consider why the effect of argument type is stronger in -wo languages. We hypothesize that this is due to a general preference for a subject question interpretation of children. The absence of evidence from word order would then cause children to be more likely to misinterpret object questions than subject questions. Secondly, we find that, only in -wo languages, reversal errors are less likely in questions with case marking on the wh phrase than in questions without case marking. This indicates that case and word order are both good cues for the children to avoid reversal errors, but that their effects are not additive, and one of the two cues suffices. Finally, consider our finding that *who*-questions are less likely than *which*-questions to receive an erroneous reverse response in the +wo languages but not in the -wo languages. This effect suggests that subjects indeed intervene for movement of *which*-phrases in +wo languages, but that in -wo languages there is a way of bypassing the intervening subject. Bypassing of the subject may be not directly related to the wo factor, but instead to word order flexibility. Specifically, several of the -wo languages (all three Slavic languages, German, Dutch; e.g. Cinque and Kayne 2005), and only one +wo language (Portuguese, Costa 2004) have been shown to allow a type movement referred to as *scrambling* that may place the object to the left of the subject even when the object is not a wh phrase. Furthermore, Greek (Alexiadou, 1999) and Italian (Burzio, 1986) exhibit word order variation similar

In the introduction we mentioned that object movement structures are frequently used for the diagnosis of SLI mostly on the basis of research on English. Do our results validate this approach? The answer is a qualified yes, but with some caution added. Our results show that at least across the languages with the structural characteristics of SOV or SVO word order and movement in questions there are some general commonalities in the acquisition profile – object questions are generally more likely to receive the reverse interpretation than subject questions. This makes it plausible that object questions or other object movement structures can be used for SLI-diagnosis in these languages. At the same time, our results show that there is also variation in the acquisition of object questions due to properties of the language and that English has at least two properties—the absence of morphological case and that it uses two different morphemes, *who* and *which*, to ask for animates—that make questions and specifically object questions difficult for a child to interpret correctly. Therefore any use of object extraction structures for diagnostic purposes needs to be accompanied by further research into the acquisition process in that language.

This type of result has great potential value to European society because object movement structures are implicated in the diagnosis and therapy of atypical child language (Adani, Forgiarini, Guasti, and Van der Lely 2014; Friedmann and Novogrodsky 2011; Levy and Friedmann 2009; van der Lely and Battell 2003; van der Lely and Pinker 2014, see also the introduction). We believe that the economic potential of first language acquisition research is widely underestimated – language acquisition research could contribute as much to European economy as the entire agricultural sector. Here is how we arrive at this estimate: Consider first that first language acquisition is crucial for later educational and economic success. In modern society, individuals with unmet speech, language, and communication needs (SLCNs) cannot contribute to their fullest potential to the overall well-being. The lost economic potential is likely enormous. For the UK, Bercow (2008) cites an estimate of the damage of unmet SLCNs in 2008 of 26 billion GBP when British GDP was 1 682 billion GBP (see also Hartshorne 2006; [Law, Zeng, Lindsay, and Beecham 2012](#), and others). The British figure represents about 1.5% of the annual gross domestic product (GDP). For the United States, Ruben (2000) provides an even higher estimate relative to GDP, namely \$154.3 to \$186 billion, i.e. 1.6 to 1.9% of 1999 GDP of \$9.661 trillion. Similar statistics are unfortunately not available for other European countries as far as we know, but we perceive the UK to be a leader for the diagnosis and therapy of SLCNs within Europe, especially because many more different diagnostic tools and procedures are available for English than for all other languages. So, an annual damage of about 1.5% GDP due to unmet SLCNs across all European countries represents in our view a conservative estimate. In countries such as Britain and Germany, this loss of GDP represents roughly 3 times the contribution the entire agricultural sector makes to GDP. For the European Union as a whole it amounts to an annual sum lost of over 210 billion euros based on the 2014 GDP figure which is roughly in line with the GDP contribution of European agriculture.

SLI is the most prevalent SLCN with 7% of an age cohort affected (Bishop, 2010). It is realistic that SLI accounts for the major share of the cost too: 1.5% of GDP would correspond to slightly more than 20% of the GDP contribution from the share of the population affected with SLI. At the same time, Bishop (2010) reports that far less research on SLI is published than on most other neurodevelopmental disorders. One factor Bishop identifies as holding back research on SLI is the lack of a critical mass of researchers. We believe one important means to increase the critical mass is to link researchers working on different languages. Indeed, the two most prevalent disorders that Bishop's analysis shows as underresearched affect primarily language, namely SLI and Speech sound disorder. Since disorders that affect language differ at least superficially across languages (Clahsen, Bartke, and Göllner 1997; Leonard 2014, and others), it is more difficult to compare research results for different languages. Part of the goal of the research reported here is to contribute to a basis for future cross-linguistic research on typical and atypical language acquisition in Europe. We think such cross-linguistic research is of growing importance, especially as migration within and to Europe contributes to the creation of multilingual societies. Finally, we hope our research facilitates practitioners spreading best practice across the boundaries between different European linguistic communities.

Notes

1. These findings contradict earlier claims by Stromswold (1995), whose work, however, was based solely on corpus data.
2. In addition, Catalan was involved at this stage of the study.
3. Italian is an exception since it is an SVO languages, but allows postverbal subjects (Burzio, 1986).
4. The object case form *whom* is archaic in English and wasn't used in this study.
5. For Standard Greek, but not Cypriot Greek, there is no case difference between the *which*-phrases in subject and object questions, i.e. Standard Greek is like German in this respect. But there is a case difference on the definite noun phrase in Standard Greek while there isn't in German. Therefore we decided to mark Standard Greek as ● rather than as ○ in table 2. But we also computed a mixed logit model like that reported below in table 3 after changing the Case value for Standard Greek complex questions to -case and found that all the predictors listed in table 3 would remain significant.
5. For predictor age, we provide as intuition of effect size the percentages correct for the two groups 4;10–5;5 vs 5;6–6;0. The log odds coefficient indicates that with each month of age the likelihood of a correct response increases by 4.89%.
7. The ratio reported excludes subject questions since clitic doubling only occurred with object questions in our sample.
8. For German, the factor matches on half of the items.
9. The gender specification of *who* varies across languages. In English *who* is genderless as it can bind either feminine and masculine pronouns (*Who_i lost his_i/her_i hat?*), while in German it can bind only a masculine pronoun (*Wer_i hat*

*seine_i/*ihre_i Mütze verloren?*). But there is no data available on this variation for most of the other languages we studied, so we leave this topic for future research. Since Adani et al. (2010) find that the effect of gender is weaker than that of case, our main findings are unlikely to be affected by gender.

10. Specifically, the test for Hebrew showed a p -value of $p = .79$ for the interaction. The lowest p -value was achieved for Portuguese ($p = .054$).
11. We also computed a χ^2 -test for the number of correct responses, which we estimated from the percent values of Friedmann et al.'s table 6, and this indicated an even greater risk of a false positive result.
12. We didn't test for any higher order interactions beyond pair-wise ones.

Acknowledgements

We dedicate this paper to the memory of Heather K. J. van der Lely. We would much prefer if she could have still been the lead author of the paper: She designed the study and as vice-chair of COST action A33 led its execution up to a presentation of preliminary results at the 2010 final conference of the action in London. We remain grateful to Heather for this and the many other contributions she made to all of our careers and lives. We thank also Paul Richens, who made relevant files from Heather's computer available to us.

We thank the many researchers and research assistants who provided technical assistance to this project, namely Sanne Berends (NL), Mechthild Bernhard (DE), Joana Cerejeira (PT), Ane Knüppel (DK), Franziska Krüger (DE), Maria Lobo (PT), Vasiliki Marcopoulou (GR), Michaela Nerantzini (GR), Lisa Raitzel (DE), Stavroula Stavrakaki (GR), Alma Veenstra (NL), and John Weston (UK). We are especially grateful to the many children who participated in this research and their parents and care-givers. Finally, we thank the audiences at the 2015 GALA conference, the University of Leiden, and Boğaziçi University, the two anonymous reviewers, Jesse Snedeker, and the two editors, Kevin Durkin and Chloe Marshall, for their many helpful suggestions.

Funding

We thankfully acknowledge the funding for COST action A33 (chair: U. Sauerland, vice-chair H. K. J. van der Lely) that made it possible to coordinate the research reported here. Furthermore the research was supported in part by the following grants: UK, IT, DE, RO, LT, and BE: European Union, CLAD 135295-LLP-2007-UK-KA1SCR; NL, DE, IT: Netherlands Organisation for Scientific Research (NWO), grant 236-78-001; AT: Austrian Science Fund (FWF), grant P 20464-G15; DE: German Research Council (DFG), grant SA 925/1-4; Bundesministerium für Bildung und Forschung (BMBF), grant 01UG1411; FI: Academy of Finland; PL: Polish Ministry of Science and Higher Education, grant 809/N-COST/2010/0, National Science Centre, grant BST no. 1445/2009; and RS: Serbian Ministry for Education, Science and Technological Development, Project ON179033.

References

- Adani, F., Forgiarini, M., Guasti, M. T., & Van der Lely, H. K. J. (2014). Number dissimilarities facilitate the comprehension of relative clauses in children with (Grammatical) Specific Language Impairment. *Journal of Child Language*, *41*, 811–841. doi: 10.1017/S0305000913000184
- Adani, F., van der Lely, H. K. J., Forgiarini, M., & Guasti, M. T. (2010). Grammatical feature dissimilarities make relative clauses easier: A comprehension study with Italian children. *Lingua*, *120*, 2148–2166. doi: 10.1016/j.lingua.2010.03.018
- Alexiadou, A. (1999). On the properties of some Greek word order patterns. In A. Alexiadou, G. Horrocks, & M. Stavrou (Eds.), *Studies in Greek syntax* (pp. 45–65). Berlin, Germany: Springer. doi: 10.1007/978-94-015-9177-5_3
- Armon-Lotem, S., Haman, E., de Lopez, K. J., Smoczynska, M., Yatsushiro, K., Szczerbinski, M., . . . van der Lely, H. (2015). A large-scale crosslinguistic investigation of the acquisition of passive. *Language Acquisition*, advance online publication. doi: 10.1080/10489223.2015.1047095
- Arosio, F., Guasti, M. T., & Stucchi, N. (2011). Disambiguating information and memory resources in children’s processing of Italian relative clauses. *Journal of Psycholinguistic Research*, *40*, 137–154. doi: 10.1007/s10936-010-9160-0
- Arosio, F., Yatsushiro, K., Forgiarini, M., & Guasti, M. T. (2011). Morphological information and memory resources in children’s processing of relative clauses in German. *Language Learning and Development*, *8*, 340–364. doi: 10.1080/15475441.2011.634691
- Avrutin, S. (2000). Comprehension of discourse-linked and non-discourse-linked questions by children and Broca’s aphasics. In Y. Grodzinsky, L. Shapiro, & D. Swinney (Eds.), *Language and the brain: Representation and processing* (pp. 295–313). San Diego, CA: Academic Press. doi: 10.1016/B978-012304260-6/50017-7
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., . . . Grothendieck, G. (2015). *lme4: Linear mixed-effects models using Eigen and S4, version 1.1-10*. R Foundation for Statistical Computing.
- Bercow, J. (2008). *Bercow review of services for children and young people (0-19) with speech, language and communication needs: Interim report*. London, UK: Department for Children, Schools and Families.
- Bishop, D. V. M. (2003). *Test for reception of grammar (TROG-2)*. Oxford, UK: Pearson.
- Bishop, D. V. M. (2010). Which neurodevelopmental disorders get researched and why? *PLoS ONE*, *5*(11), e15112. doi: 10.1371/journal.pone.0015112
- Burzio, L. (1986). *Italian syntax: A government-binding approach*. Dordrecht, Netherlands: Kluwer.
- Chomsky, N. (1957). *Syntactic structures*. The Hague, Netherlands: Mouton.
- Chomsky, N. (1981). *Lectures on government and binding: The Pisa lectures*.

- Berlin, Germany: Mouton de Gruyter.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Cinque, G., & Kayne, R. S. (Eds.). (2005). *The Oxford handbook of comparative syntax*. Oxford, UK: Oxford University Press. doi: 10.1093/oxfordhb/9780195136517.001.0001
- Clahsen, H., Bartke, S., & Göllner, S. (1997). Formal features in impaired grammars: A comparison of English and German SLI children. *Journal of Neurolinguistics*, 10, 151–171. doi: 10.1016/S0911-6044(97)00006-7
- Costa, J. (2004). *Subject positions and interfaces: the case of European Portuguese*. Berlin, Germany: Walter de Gruyter.
- Friedmann, N., Belletti, A., & Rizzi, L. (2009). Relativized relatives: Types of intervention in the acquisition of A-bar dependencies. *Lingua*, 119, 67–88. doi: 10.1016/j.lingua.2008.09.002
- Friedmann, N., & Novogrodsky, R. (2011). Which questions are most difficult to understand?: The comprehension of wh questions in three subtypes of SLI. *Lingua*, 121, 367–382. doi: 10.1016/j.lingua.2010.10.004
- Gardner, H., Froud, K., McClelland, A., & van der Lely, H. (2006). Development of the grammar and phonology screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language & Communication Disorders*, 41, 513–540. doi: 10.1080/13682820500442644
- Goodluck, H. (2010). Object extraction is not subject to child relativized minimality. *Lingua*, 120, 1516–1521. (Contrast as an information-structural notion in grammar) doi: DOI: 10.1016/j.lingua.2009.10.005
- Grillo, N. (2009). Generalized minimality: Feature impoverishment and comprehension deficits in agrammatism. *Lingua*, 119, 1426–1443. doi: 10.1016/j.lingua.2008.04.003
- Guasti, M. T., Branchini, C., & Arosio, F. (2012). Interference in the production of Italian subject and object wh-questions. *Applied Psycholinguistics*, 33, 185–223. doi: 10.1017/S0142716411000324
- Guasti, M. T., Stavrakaki, S., & Arosio, F. (2012). Cross-linguistic differences and similarities in the acquisition of relative clauses: Evidence from Greek and Italian. *Lingua*, 122, 700–713. doi: 10.1016/j.lingua.2012.02.001
- Hartshorne, M. (2006). *The cost to the nation of children's poor communication*. London, UK: I CAN (www.ican.org.uk).
- Hosmer Jr., D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: John Wiley & Sons. doi: 10.1002/9781118548387
- Jaeger, F. T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. doi: 10.1016/j.jml.2007.11.007
- Johnson, V. E., de Villiers, J. G., & Seymour, H. N. (2005). Agreement without understanding? The case of third person singular /s/. *First Language*, 25, 317–330. doi: 10.1177/0142723705053120
- Katsos, N., Ezeizabarrena, M.-J., Gavarró, A., Kuvac Kraljevic, J., Hrzica, G., Grohmann, K. K., . . . Noveck, I. (2012). The acquisition of quantification

- across languages: some predictions. In *Proceedings of the 36th annual Boston University conference on language development* (pp. 258–268). Somerville, MA: Cascadilla Press.
- Law, J., Zeng, B., Lindsay, G., & Beecham, J. (2012). Cost-effectiveness of interventions for children with speech, language and communication needs (SLCN): a review using the Drummond and Jefferson (1996) 'referee's checklist'. *International Journal of Language & Communication Disorders*, 47, 1–10. doi: 10.1111/j.1460-6984.2011.00084.x
- Leonard, L. B. (2014). Specific language impairment across languages. *Child Development Perspectives*, 8, 1–5. doi: 10.1111/cdep.12053
- Levy, H., & Friedmann, N. (2009). Treatment of syntactic movement in syntactic SLI: A case study. *First Language*, 29, 15–49. doi: 10.1177/0142723708097815
- Metz, M., van Hout, A., & Lely, H. v. d. (2012). Subject interpretation of object questions by dutch 5-year-olds: The role of number agreement in comprehension. In *Linguistics in the netherlands* (Vol. 29, pp. 97–110). Amsterdam, Netherlands: John Benjamins Publishing Company. doi: 10.1075/avt.29.08met
- Metz, M., van Hout, A., & van der Lely, H. (2010). Understanding who and which questions in five to nine-year-old Dutch children: the role of number. *GAGL: Groninger Arbeiten zur germanistischen Linguistik*, 51, 27–41. Retrieved from <http://irs.ub.rug.nl/dbi/4ef4a1ef1c357>
- Pesetsky, D. (1987). *Wh-in-situ*: Movement and unselective binding. In E. Reuland & A. ter Meulen (Eds.), *The representation of (in)definiteness* (p. 98–129). Cambridge, MA: MIT Press.
- R Core Team. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. R Foundation for Statistical Computing.
- Rizzi, L. (1990). *Relativized minimality*. Cambridge, MA: MIT Press. doi: 10.9793/elsj1984.8.154
- Rizzi, L. (2004). Locality and the left periphery. In A. Belletti (Ed.), *Structures and beyond: The cartography of syntactic structures* (pp. 223–251). Oxford University Press.
- Rizzi, L. (2013). Locality. *Lingua*, 130, 169–186. doi: 10.1016/j.lingua.2012.12.002
- Ruben, R. J. (2000). Redefining the survival of the fittest: communication disorders in the 21st century. *The Laryngoscope*, 110, 241–241. doi: 10.1097/00005537-200002010-00010
- Sauerland, U., Anderssen, J., & Yatsushiro, K. (2005). The plural is semantically unmarked. In S. Kepser & M. Reis (Eds.), *Linguistic evidence – empirical, theoretical, and computational perspectives*. Berlin, Germany: Mouton de Gruyter. doi: 10.1515/9783110197549.413
- Schouwenaars, A., van Hout, A., & Hendriks, P. (2012). Word order overrules number agreement: Dutch children's interpretation and production of *which*-questions. In C.-Y. Chu et al. (Eds.), *Selected proceedings of the 5th GALANA conference* (p. 60–71). Somerville, MA: Cascadilla.

- Seidl, A., Hollich, G., & Jusczyk, P. W. (2003). Early understanding of subject and object wh-questions. *Infancy, 4*, 423–436. doi: [10.1207/S15327078IN0403_06](https://doi.org/10.1207/S15327078IN0403_06)
- Strangmann, I., Slomp, A., & van Hout, A. (2014). Development of Dutch children's comprehension of subject and object wh-questions: The role of topicality. *Linguistics in the Netherlands, 31*, 129–144. doi: [10.1075/avt.31.10str](https://doi.org/10.1075/avt.31.10str)
- Stromswold, K. (1995). The acquisition of subject and object wh-questions. *Language Acquisition, 4*, 5–48. doi: [10.1080/10489223.1995.9671658](https://doi.org/10.1080/10489223.1995.9671658)
- Thornton, R. J. (1990). *Adventures in long-distance moving: The acquisition of complex wh-questions* (Unpublished doctoral dissertation). University of Connecticut, Storrs.
- Tieu, L., Bill, C., Romoli, J., & Crain, S. (2014). Plurality inferences are scalar implicatures: Evidence from acquisition. In *Proceedings of SALT 24* (pp. 122–136). Washington, DC: Linguistic Society of America. doi: [10.3765/salt.v24i0.2421](https://doi.org/10.3765/salt.v24i0.2421)
- Tribushinina, E., van den Bergh, H., Kilani-Schoch, M., Aksu-Koç, A., Dabašinskienė, I., Hrzica, G., ... Dressler, W. (2013). The role of explicit contrast in adjective acquisition: A cross-linguistic longitudinal study of adjective production in spontaneous child speech and parental input. *First Language, 33*, 594–616. doi: [10.1177/0142723713503146](https://doi.org/10.1177/0142723713503146)
- van der Lely, H. K. J., & Battell, J. (2003). Wh-movement in children with grammatical SLI: A test of the RDDR hypothesis. *Language, 79*, 153–181. doi: [10.1353/lan.2003.0089](https://doi.org/10.1353/lan.2003.0089)
- van der Lely, H. K. J., Gardner, H., McClelland, A. G. R., & Froud, K. E. (2007). *Grammar and phonology screening test (GAPS)*. London, UK: DLDCN.com.
- van der Lely, H. K. J., & Pinker, S. (2014). The biological basis of language: insight from developmental grammatical impairments. *Trends in Cognitive Sciences, 18*, 586–595. doi: [10.1016/j.tics.2014.07.001](https://doi.org/10.1016/j.tics.2014.07.001)
- Varlokosta, S., Belletti, A., Costa, J., Friedmann, N., Gavarró, A., Grohmann, K. K., ... Yatsushiro, K. (2015). A cross-linguistic study of the acquisition of clitic and pronoun production. *Language Acquisition*, advance online publication. doi: [10.1080/10489223.2015.1028628](https://doi.org/10.1080/10489223.2015.1028628)
- Varlokosta, S., Nerantzini, M., & Papadopoulou, D. (2015). Comprehension asymmetries in language acquisition: a test for relativized minimality. *Journal of Child Language, 42*, 618–661. doi: [10.1017/S0305000914000257](https://doi.org/10.1017/S0305000914000257)
- Varnava, M., & Grohmann, K. K. (2014). Developments in the acquisition of wh-interrogatives in Cypriot Greek. *Linguistic Variation Yearbook, 14*, 69–108. doi: [10.1075/lv.14.1.04var](https://doi.org/10.1075/lv.14.1.04var)
- Xanthos, A., Laaha, S., Gillis, S., Stephany, U., Aksu-Koç, A., Christofidou, A., ... Dressler, W. U. (2011). On the role of morphological richness in the early development of noun and verb inflection. *First Language, 31*, 461–479. doi: [10.1177/0142723711409976](https://doi.org/10.1177/0142723711409976)

All 24 original items

- (1) Who is pushing the grannies?
- (2) Who is pulling the ladies?
- (3) Who is washing the dancers?
- (4) Who is catching the princesses? (excluded)
- (5) Who is scratching the queens?
- (6) Who is feeding the fairies?
- (7) Which fairy is pushing the queens?
- (8) Which queen is pulling the princesses?
- (9) Which princess is washing the dancers?
- (10) Which dancer is catching the ladies? (excluded)
- (11) Which lady is scratching the grannies?
- (12) Which grannie is feeding the fairies?
- (13) Who are the dancers pushing? (excluded)
- (14) Who are the princesses pulling?
- (15) Who are the queens washing?
- (16) Who are the fairies catching? (excluded)
- (17) Who are the grannies scratching?
- (18) Who are the ladies feeding?
- (19) Which princess are the dancers pushing?
- (20) Which dancer are the ladies pulling?
- (21) Which lady are the grannies washing?
- (22) Which grannie are the fairies catching? (excluded)
- (23) Which fairy are the queens scratching?
- (24) Which queen are the princesses feeding?

Glosses of Sample Items

Abbreviations used: SG - singular, PL - plural, NOM - nominative case, PART - partitive case, ACC - accusative case, PROG - progressive, DEF - definite, M - masculine, F - feminine, 3 - third person, CL - clitic.

EE: Estonian

- (1) a. Kes lükka-b printsesse?
who.NOM push-3SG princess.PL.PART
'Who is pushing the princesses?'
- b. Missugune naine lükka-b kuninganna-sid?
which.NOM lady.NOM push-3SG queen-PL.PART

- ‘Which lady is pushing the queens?’*
- c. Ke-da printsessi-d lükka-vad?
 who-PART princess-PL.NOM push-3PL
‘Who are the princesses pushing?’
- d. Missugus-t printsessi naise-d lükka-vad?
 which-PART princess.PART lady-PL.NOM push-3PL
‘Which princess are the ladies pushing?’

FI: Finnish

- (2) a. Kuka työntä-ä mummo-j-a?
 who push-3SG granny-PL-PART
‘Who is pushing the grannies?’
- b. Kuka keiju työntä-ä kuningattar-i-a?
 who fairy.SG push-3SG queen-PL-PART
‘Which fairy is pushing the queens?’
- c. Ketä tanssija-t työntä-vät?
 who.PART dancer-PL push-3PL
‘Who are the dancers pushing?’
- d. Ketä prinsessa-a tanssija-t työntä-vät
 who-PART princess-PART dancer-PL push-3PL
‘Which princess are the dancers pushing?’

MT: Maltese

- (3) a. Min qed jimbotta lin-nanniet?
 who PROG push.3M.SG to.the-grannies
‘Who is pushing the grannies?’
- b. Liema ballerina qed taqbad lin-nisa?
 which dancer PROG catch.3F.SG to.the-ladies
‘Which dancer is catching the ladies?’
- c. Lil min qed jimbuttaw il-ballerini?
 to who PROG push-3PL the-dancers
‘Who are the dancers pushing?’
- d. Liema prinçipessa qed jimbuttaw il-ballerini?
 which princess PROG push-3PL the-dancers?
‘Which princess are the dancers pushing?’

IL: Hebrew

- (4) a. Mi doxefet et ha-nesixot?
 who push.SG-FM ACC DEF-princess.PL
‘Who pushes the princesses?’
- b. Eizo isha doxefet et ha-malkot?
 which.FM woman push.SG-F ACC DEF-queen.PL
‘Which woman pushes the queens?’

- c. Et mi ha-nesixot doxafot?
ACC who DEF-princess.PL push.PL.FM
'Who are the princesses pushing?'
- d. Et eizo nesixa ha-nashim doxafot?
ACC which.FM princess DEF-woman.PL push.PL.FM
'Which woman are the princesses pushing?'

EL: Greek

- (5) a. Pjos sproxni tis jajades?
who.M.NOM push.3SG the.F.PL.ACC granny.F.PL.ACC
'Who is pushing the grannies?'
- b. Pja neraida sproxni tis vasilises?
who.F fairy.F push.3SG the.F.PL.ACC queen.F.PL.ACC
'Which fairy is pushing the queens?'
- c. Pjon sproxnun i horeftries?
who.M.ACC push.3SG the.F.PL.NOM dancer.F.PL.NOM
'Who are the dancers pushing?'
- d. Pja prigipisa sproxnun i horeftries?
who.F princess.F push.3SG the.F.PL.NOM dancer.F.PL.NOM
'Which princess are the dancers pushing?'

CY: Cypriot Greek

- (6) a. Pjos kounta tes jiajiaes?
who.M.NOM push.3SG the.PL.ACC granny.PL.ACC
'Who is pushing the grannies?'
- b. Pja neraida kounta tes vasilisses?
who.F.NOM fairy.NOM push.3SG the.PL.ACC queen.PL.ACC
'Which fairy is pushing the queens?'
- c. Pjon kountoun i horeftries?
who.M.ACC push.3PL the.PL.NOM dancer.PL.NOM
'Who are the dancers pushing?'
- d. Pjan pringipissan kountoun i horeftries?
who.F.ACC princess.ACC push.3PL the.PL.NOM
dancer.PL.NOM
'Which princess are the dancers pushing?'

LT: Lithuanian

- (7) a. Kas traukia moteris?
who.NOM pull.3SG lady.PL.ACC
'Who is pulling the ladies?'
- b. Kuri balerina gaudo moteris?
which.SG.F.NOM dancer.SG.NOM catch.3SG lady.PL.ACC

- ‘Which dancer is catching the ladies?’*
- c. Ką traukia princesės?
who.ACC pull.3PL princess.PL.NOM
‘Who are the princesses pulling?’
- d. Kurią princesę stumia balerinos?
which.SG.F.ACC princess.SG.ACC push.3PL dancer.PL.NOM?
‘Which princess are the dancers pushing?’

HR: Croatian

- (8) a. Tko vuče dame?
who.NOM pull.3SG lady.PL.ACC
‘Who pulls the ladies?’
- b. Koja balerina lovi dame?
which.SG.F.NOM dancer.NOM catch.3SG lady.PL.ACC
‘Which dancer catches the ladies?’
- c. Koga vuku princeze?
who.ACC pull.3PL princess.PL.NOM
‘Who do the princesses pull?’
- d. Koju princezu guraju balerine?
which.SG.F.ACC princess.SG.ACC push.3PL dancer.PL.NOM
‘Which princess do the dancers push?’

RS: Serbian

- (9) a. Ko gura princeze?
who.NOM push.3SG princess.ACC.PL.F
‘Who pushes the princesses?’
- b. Koja žena gura kraljice?
which.NOM.SG.F lady.NOM.SG.F push.3SG queen.ACC.PL.F
‘Which lady pushes the queens?’
- c. Koga guraju princeze?
who.ACC push.3PL princess.NOM.PL.F
‘Who do the princesses push?’
- d. Koju princezu guraju žene?
which.ACC.SG.F princess.ACC.SG.F push.3PL lady.NOM.PL.F
‘Which princess do the ladies push?’

PL: Polish

- (10) a. Kto ciągnie panie?
who.NOM pull.3SG lady.PL.ACC
‘Who pulls the ladies?’
- b. Która tancerka łapie panie?
which.SG.F.NOM dancer.NOM catch.3SG lady.PL.ACC
‘Which dancer catches the ladies?’

- c. Kogo ciągną księżniczki?
Who.SG.ACC pull.3PL princess.PL.NOM
'Who do the princesses pull?'
- d. Którą księżniczkę popychają tancerki?
which.SG.F.ACC princess.SG.ACC push.3PL dancer.PL.NOM
'Which princess do the dancers push?'

UK: English

- (11) a. Who is pushing the princess?
b. Which lady is pushing the queen
c. Who are the princesses pushing?
d. Which princess are the ladies pushing

DE: German & AT: Austrian German

- (12) a. Wer zieht die Frauen?
who.NOM pull.3.SG the.PL woman.PL.ACC
'Who pulls the women?'
- b. Welche Tänzerin fängt die Frauen?
which.FEM dancer.SG.NOM catch.3.SG the.PL woman.PL.ACC
'Which dancer catches the women?'
- c. Wen ziehen die Prinzessinnen?
who.ACC pull.PL the.PL princess.PL.NOM
'Who do the princesses pull?'
- d. Welche Prinzessin schubsen die Tänzerinnen?
which.FEM princess.SG.ACC push.3.PL the.PL dancer.PL.NOM
'Which princess do the dancers push?'

NL: Dutch

- (13) a. Wie trekt de vrouwen?
who push.3SG the lady.PL
'Who pushes the ladies?'
- b. Welke danseres vangt de vrouwen?
which dancer catch.3SG the lady.PL
'Which dancer catches the ladies?'
- c. Wie trekken de prinsessen?
who pull.3PL the princess.PL
'Who pulls the princesses?' / 'Who do the princesses pull?'
- d. Welke prinses duwen de danseressen?
which princess push.3PL the dancer.PL
'Which princess do the dancers push?'

DK: Danish

- (14) a. Hvem er ved at trække damerne?
who is by to pull lady.PL.DEF

- 'Who is pulling the ladies?'*
- b. Hvilken danser er ved at fange damerne?
which dancer.SG is by to catch lady.PL.DEF
'Which dancer is pulling the ladies?'
- c. Hvem er prinsesserne ved at trække?
who is princess.PL.DEF by to push
'Who are the princesses pushing?'
- d. Hvilken prinsesse er danserne ved at skubbe?
which prinsesse.SG is dancer.PL.DEF by to push
'Which princess are the dancers pushing?'

BE: French

- (15) a. Qui est ce qui gratte les reines?
who is it who scratch the queen.PL
'Who scratches the queens?'
- b. Quelle est la danseuse qui attrape les femmes?
which is the dancer who catch the woman.PL
'Which dancer catches the women?'
- c. Qui les grands-mères grattent-elles?
who the grand-mother.PL scratch-they
'Who do the grannies scratch?'
- d. Quelle reine les princesses nourrissent-elles?
which queen the princess.PL feeding-they
'Which queen do the princesses feed?'

IT: Italian

- (16) a. Chi spinge le principesse?
who push.3SG the.PL princess.PL
'Who pushes the princesses?'
- b. Quale signora spinge le regine?
which lady push.3SG the.PL queen.PL
'Which lady pushes the queens?'
- c. Chi spingono le principesse?
who push.PL the princess.PL
'Who do the princesses push?'
- d. Quale principessa spingono le signore?
which princess push.PL the.PL lady.PL
'Which princess do the ladies push?'

PT: European Portuguese

- (17) a. Quem é que está a puxar as senhoras?
who is.3SG that at is.3SG pulling the.PL lady.PL
'Who is pulling the ladies?'

- b. Que bailarina é que está a apanhar as senhoras?
 which dancer is.3SG that is.3SG at catching the.PL lady.PL
 ‘Which dancer is catching the ladies?’
- c. Quem é que as princesas estão a puxar?
 who is.3G that the.PL princess.PL are.3PL at pulling
 ‘Who are the princesses pulling?’
- d. Que princesa é que as bailarinas estão a empurrar?
 which princess is.3SG that the.PL dancer.PL are.3PL at pushing
 ‘Which princess are the dancers pushing?’

RO: Romanian

- (18) a. Cine trage doamnele?
 who.NOM pull.3SG ladies.the
 ‘Who pulls the ladies?’
- b. Care regină trage prințesele?
 which.NOM queen pull.3SG princesses.the
 ‘Which queen pulls the princesses?’
- c. Pe cine trag prințesele?
 ACC who pull.3PL princesses.the
 ‘Who do the princesses pull?’
- d. Pe care balerină o trag doamnele?
 ACC which dancer.F.SG her.CL.ACC.F.SG pull.3PL ladies.the
 ‘Which dancer do the ladies pull?’