The ITB Journal

Volume 13 | Issue 1 Article 2

2012

Wh-Interrogatives in Early L1 Greek: Comprehension vs. Production

Evangelia K. Asproudi

Wh-Interrogatives in Early L1 Greek: Comprehension vs. Production

Evangelia K. Asproudi
School of English
Aristotle University of Thessaloniki, Greece

Abstract

This paper examines the comprehension and production of wh-interrogatives in early L1 Greek. Specifically, children's performance is explored at both levels with regard to argument/adjunct extraction and presence/absence of negation. In order to test the predicted lead of comprehension over production, a group of ninety four-to-seven-year old Greek children participated in elicitation tasks that were designed mainly along the methodological principles of Crain and Thornton (1998). On the whole, the results were in line with the initial expectation. That is, the Greek children performed better in question comprehension compared to question production, with individual findings suggesting that children's economy-based processing may not be constrained exclusively by syntactic factors.

Keywords: *wh-interrogatives, comprehension, production, L1 Greek, extraction site, negation*

1. Introduction

The aim of this paper is to investigate children's behaviour with regard to the acquisition of wh-questions in L1 Greek. More concretely, question comprehension and production will be examined in an attempt to provide a thorough picture of what happens with the acquisition of wh-movement in the language.

The core hypothesis underlying this paper is that in languages clause-typing wh-questions by overt wh-movement (Cheng 1991, 1997), children acquire wh-movement from a very young age. Short-distance (henceforth SD) movement is present in child grammar from very early on (e.g. Guasti 2000 etc.), while access to long-distance (henceforth LD) movement occurs from around the age of three onwards (e.g. Thornton 1990, Thornton & Crain 1994, de Villiers et al. 1990 etc.). At this point, a distinction should be made with regard to children's performance on comprehension and production of wh-movement. Specifically, it is hypothesized that children perform better at the comprehension than at the production level, given than the latter involves more processing demands than the former.

In view of the above hypotheses, the prediction with regard to the acquisition of wh-questions in early L1 Greek is that Greek children will perform better at the comprehension than the production level, with accurate performance rates increasing with age.

2. Method

In order to test the above prediction concerning early L1 Greek, a study was designed in which the methodology of research adopted was one of elicited comprehension and of elicited production. What follows is a description of the participating subjects, of the

experimental tasks set and of the procedure followed, as well as of the measurement steps employed for the analysis of the collected data.

2.1 Participants

The study group consisted of ninety typically developing children aged 4;0 to 7;0. These children were divided into three equivalent subgroups A, B and C. Group A included thirty children between four and five (mean age range: 4;6), group B thirty children between five and six (mean age range: 5;5) and group C thirty children between six and seven years old (mean age range: 6;7). Group A and B children were in their first and second year in kindergarten respectively, while group C children attended the first grade in primary school.

2.2 Procedure and Materials

As stated earlier, the present study aims to investigate areas of L1 question comprehension and production. For this purpose, three experimental tasks were prepared for data collection. These tasks were designed to investigate the research areas as follows:

- Task 1: *comprehension of questions without wh-islands (wh-COMP*[1] *questions) and of questions with wh-islands (wh-wh questions)*
- Task 2: *production of LD questions*
- Task 3: *production of SD questions*

These experimental tasks will be described in more detail in the following sections through the presentation of the rationale, the materials and the procedure pertaining to each one of them.

Before turning to the presentation of each experimental task in isolation, some general procedural remarks applying to all tasks set should be mentioned. As regards the setting, each child was tested separately in a room next to their classroom during their daily school program. The testing of each child involved two sessions that took place on different days; each session lasted about forty-five minutes, and it was tape recorded and transcribed at a later stage. Tasks 1 and 2 were conducted during the first session, while task 3 during the second one.

What follows is a detailed presentation of the three experimental tasks mentioned above.

Task 1: Comprehension of wh-COMP and wh-wh Questions

Rationale. The rationale behind task 1 was twofold. On the one hand, the aim was to see whether children would prefer SD over LD movement at the comprehension level, and for this reason ARG-COMP and ADJ-COMP[2] questions were included where both a SD and a LD interpretation was grammatical and plausible. On the other hand, there was also an interest in whether children would show sensitivity to wh-islands. In order to test this, a set of argument- and adjunct- medial questions was also included that permitted the children a choice between a grammatical SD and an ungrammatical LD

[1] COMP refers throughout to an overt non-wh-complementiser.
[2] ARG = argument, ADJ = adjunct

interpretation. On the whole, the expectations were that children would show increasing preference for SD interpretations and sensitivity to island constraints.

Materials. For the purpose of this task, short stories were created which replicated scenarios used in similar studies in other languages (e.g. Abdulkarim et al. 1997, Philip & de Villiers 1992, Roeper & de Villiers 1992, Thornton & Crain 1994, de Villiers et al. 1990), or which were adapted from stories included in school text books in an attempt to create materials that would not be remote from children's school experience. All stories were followed by comprehension questions; each story provided a context that made all interpretations (grammatical and ungrammatical ones) salient and that excluded any possibility of coreferentiality of SD and LD interpretations. To preclude a bias for one interpretation over the other, care was taken to deliver all test questions in as neutral intonation as possible. A sample test story is presented below.

(*English translation*)
The dog has a ball. The cat and the rabbit have to climb up a wall to see who has the ball. The cat tries first: she takes a ladder and tries to climb up the wall but eventually falls down. Then the rabbit tries: she uses the same ladder, climbs up the wall and sees the dog. Then she says: "I can see who has the ball! The dog has it and he is holding it with his legs!"

Experimenter's question: *Pjos$_{i/j}$* ***t$_i$*** *emathe oti* ***t$_j$*** *exi ti bala?*
'Who$_i$ **t$_i$** found out that she has the ball?' – SD reading
'Who$_i$ did she find out **t$_i$** to have the ball?' – LD reading

Target answer: *to kuneli* 'the rabbit' – SD reading
o skilos 'the dog' – LD reading

Experimenter's question: *Pws$_i$ emathe* ***t$_i$*** *pjos exi* ****t$_i$*** *ti bala?*
'How$_i$ did she find out **t$_i$** who has ****t$_i$*** the ball?'

Target answer: *skarfalonondas ton tiho* 'by climbing up the wall'

Procedure. The duration of this task was about twenty-five minutes. After an explanation of the procedure to follow, each of the stories was acted out with props in front of the child or presented through pictures to her. Then, the accompanying comprehension questions were posited. In case the child did not respond to a certain test question, this question was repeated once; if no answer was elicited, the procedure proceeded with the presentation of the next question or of the next story and its accompanying questions.

Task 2: Production of LD questions.
Rationale. In task 2 the aim was the production of LD questions by children. The rationale of this task evolved around the hypothesis that LD extraction is more processing costly for children than SD extraction, and hence difficult to produce.

Materials. For the elicitation of LD questions two games were designed along the lines of similar crosslinguistic tasks (Thornton 1996). For the purposes of this task, a puppet called 'Astrulis' that had come from another planet was introduced to the child. The first game included prompts of the type *Rotise ton Astruli X mandevi aftos* ('Ask

Astrulis [i.e. the puppet] X he guesses'), where X stood for the respective wh-element. As for the second game, it consisted of prompts which were of the form *Rotise ton Astruli X protimai aftos* ('Ask Astrulis [i.e. the puppet] X he prefers'), where X stood again for the respective wh-element.
What is presented below is a sample of the first and second game protocol as well as of the stimulus sentences used.

1st game:
The experimenter presents five toys: one car, one pink and one blue comb, and one red and one yellow ball. She asks both the child and the puppet to cover their eyes while she is hiding each of these toys in different numbered boxes. Then she asks the child and the puppet to uncover their eyes, and the game proceeds. After the experimenter has elicited the child's guess, she prompts the child to elicit the puppet's guess. One of the stimulus sentences she uses is the following:

Stimulus sentences:
Experimenter: *Sto kuti 3 ekripsa mia xtena. Ja mandepse pja.*
'In box 3 I hid one of the combs. Guess which one.'

(the child says his/her guess…)

Experimenter: *Esi mandevis oti sto kuti 3 ekripsa tin ... xtena. Rotise ton Astruli pja mandevi aftos.*
'You are guessing that in box 3 I hid the … comb. Ask Astrulis which one he is guessing.'

Target question: *Pja xtena mandevis oti ekripse sto kuti 3?*
'Which comb are you guessing that she (i.e. the experimenter) hid in box 3?'

2nd game:
The experimenter presents four toy characters: one rabbit, one dog and two horses. She explains to the child that three of these toy characters have to be matched with certain actions.

Stimulus sentences:

Experimenter: *Kapjo alogaki tha pai volta. Rotise ton Astruli pjo protimai aftos.*
'One of the two horses is going for a walk. Ask Astrulis which one he prefers.'

Target question: *Pjo alogaki protimas na pai volta?*
'Which horse do you prefer that it go for a walk?'

Procedure. The duration of task 2 was about twenty minutes, fifteen minutes for the first and five minutes for the second game. As regards the procedure, both games were based on ideas from Thornton (1996). The first game proceeded as follows: both the child and the puppet covered their eyes while the experimenter hid objects in small numbered boxes. The child guessed what was hidden in each box, and then the experimenter prompted her to ask the puppet about his guess. In the second game, the

child and the puppet saw some toy characters and some possible actions they could do. The experimenter prompted the child to ask the puppet decide which action each toy character would do. Thus, through these two games, the child was prompted to produce the target LD questions. In case the child did not react to the experimenter's prompt, the prompt (and the relevant part of the game) was repeated twice; if still no question was elicited, the procedure continued with the presentation of the next part of the game and the corresponding prompt.

Task 3: Production of SD questions.

Rationale. Task 3 aimed at the production of SD questions. The main rationale underlying this task was that SD question production is in line with local preference processing accounts (e.g. Fanselow et al. 1999, Frazier & Flores d' Arcais 1989 etc.), and hence it would not posit any serious problems to children.

Materials. The puppet called 'Astrulis' was again employed for the purposes of the present task. The prompts were designed on the basis of scenarios and short stories close to children's everyday life. All prompts ended either on the phrase *Rotise ton Astruli X* ('Ask Astrulis [i.e. the puppet] X'), where X stood each time for the respective wh-element, or on the phrase *Rotise ton Y X* ('Ask Y X'), where X stood each time for the respective wh-element and Y for a person from the acted-out story. Here is a sample set of the test stories and of the relevant stimulus sentences:

1st situation:

(English translation)

Three smurfs are roller-skating in the forest when they meet two horses. While playing with the two horses, a bear suddenly appears and treads down the brown horse.

Stimulus sentences:

Experimenter: *I arkuda patise ena alogaki. Rotise ton Astruli pjo.*

'The bear trod down one of the horses. Ask Astrulis which one.'

Target question: *Pjo alogaki patise i arkuda?*

'Which horse did the bear tread down?'

2nd situation:

(English translation)

Jim goes shopping. He wants to buy a ball. He goes in a shop and sees two balls, a red one and a yellow one.

Stimulus sentences:

Experimenter: *O politis lei ston Dimitri na min pari kapja bala. Rotise ton pja.*

'The salesman tells Jim not to take one specific ball. Ask him which one.'

Target question: *Pja bala na min pari o Dimitris?;*

'Which ball should Jim not take?'

Procedure. The duration of the present task was about forty-five minutes. In the first part of this task, the experimenter prompted the child to pose questions to the puppet about himself and his friends and about life on his planet (e.g. Thornton 1996). Then, both the child and the puppet saw stories acted out with props in front of them or

accompanied by pictures. At the end of each story, using lead-in statements the experimenter prompted the child to ask the puppet the target questions so as to see whether the puppet understood what had happened in the story. In the second part of the task, a big story was acted out in front of the child and the puppet, with the latter playing this time the role of a passive observer. At frequent intervals, the child had to pose questions to the toy characters of the story, so as to find out how the story continued. In cases that the child did not react to the experimenter's prompt, the prompt (and the story where necessary) was repeated twice; if still no question was elicited, the procedure continued with the presentation of the next story and its accompanying prompt.

2.3 Analyses and measurements

The data collected from all three experiments were inserted into a database using the statistical software SPSS 17.0 for Windows. Mean ratings were then calculated for each participant; these ratings were tested statistically with mixed ANOVA analyses which provided within- and between- subject comparisons. These in turn enabled the checking for significance of the effect of category and group on the ratings. On the whole, the analyses of the data carried out were the following: a. correct responses with regard to the prediction tested, and b. type of errors.

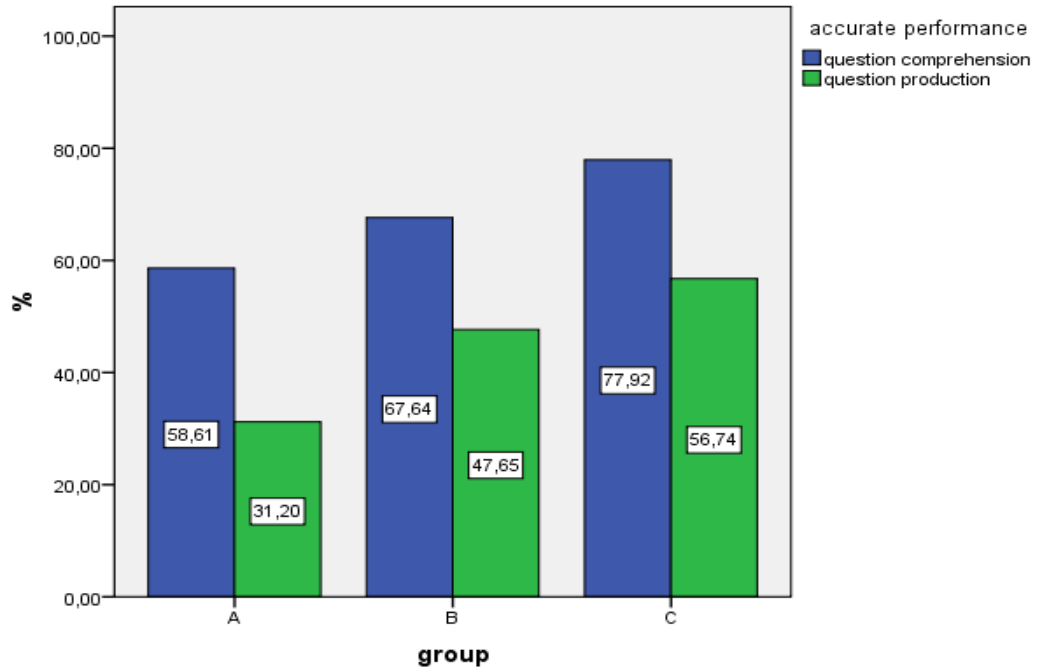
3. Results

Overall, the results obtained from tasks 1, 2 and 3 were in line with the prediction formulated in the introductory section[3]. As a reminder, the prediction was that children's performance will become more and more accurate as age increases and that children will show higher accurate performance on question comprehension than on question production overall. The following figure provides the average accurate performance rates of all groups. It should be noted that the phrase 'accurate performance' is used in a very general sense here and subsumes under it various responses on the part of the children. In terms of comprehension, 'accurate performance' includes grammatical SD/LD readings of ambiguous and unambiguous questions, namely of questions that allow either both readings or only one reading respectively. In terms of production, accurate performance implies production of target questions with respect to SD/LD movement and wh production. Separate results for all types of questions tested in comprehension and production will be provided in figures 2-5. To begin with, here is a presentation of the overall accurate performance rates attested in question comprehension and production.

As shown in Figure 1, for both conditions accuracy rates increased with age. In addition, within-group differences are also apparent, with accurate comprehension rates being higher than accurate production ones within all groups.

[3] It should be noted that performance and every elicitation on the part of the children, either at the comprehension or the production level, will be characterised throughout in terms of 'accuracy'. The term 'target' will refer to the responses/structures aimed at by the experimenter.

Figure 1. Overall Accurate Performance in Question Comprehension vs. Question Production



A two-way mixed ANOVA was performed on the results; it revealed significant within-group effects of accurate performance [F(1,87)=105.641, p=.000]. No significant interaction was found between accurate performance and group [F(2,87)=1.127, p=.329] but only a significant between-group effect [F(2,87)=16.949, p=.000], which suggests that there is large heterogeneity in performance between groups. Pairwise comparisons showed significant within-group differences in all groups (p=.000), which indicate that the Greek children performed significantly better at the comprehension than the production level from the youngest age. As for between-group comparisons, they were all found to be significantly different in question comprehension (A-B: p=.053, A-C: p=.000, B-C=.01), which suggests a clear pattern of development in children's comprehension ability. In question production, on the other hand, a burst in accurate performance was attested by group B, with A-B (p=.002) and A-C (p=.000) comparisons being the significantly different ones.

Analyzing the above data in more detail, it is interesting to see how children performed with respect to the comprehension and production of specific types of questions. On the comprehension side, two types of questions were tested: argument and adjunct questions with no wh-island (wh-COMP), and argument and adjunct questions involving a wh-island (wh-wh). The accurate performance rates on these two types of questions are shown in figures 2 and 3 respectively.

Figure 2. Accurate Performance in the Comprehension of Wh-COMP Questions

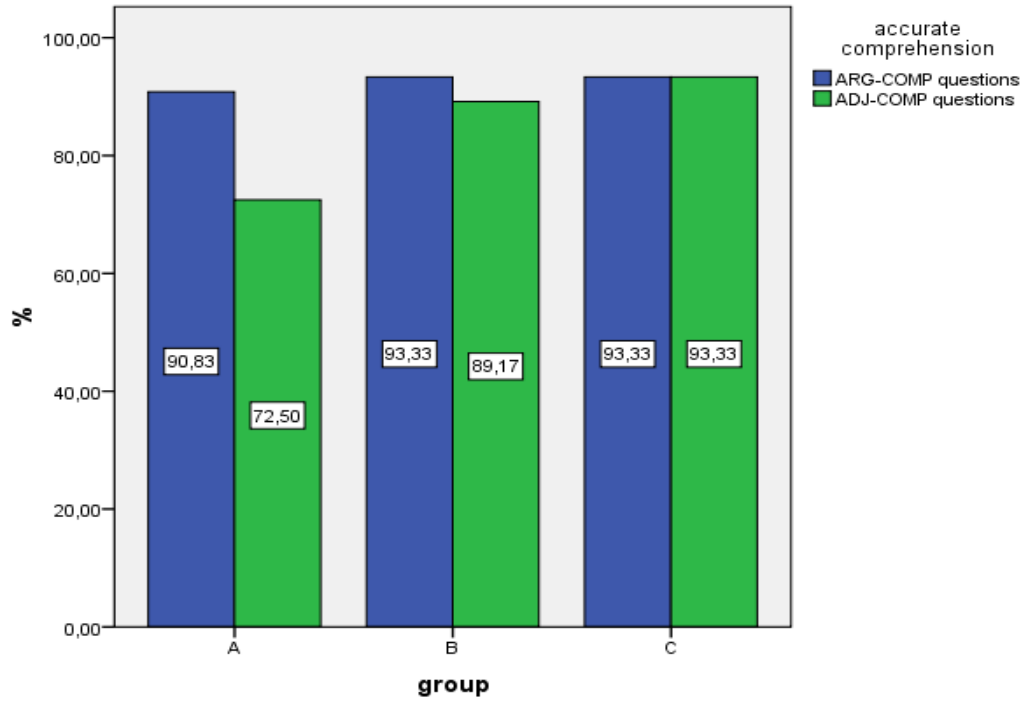
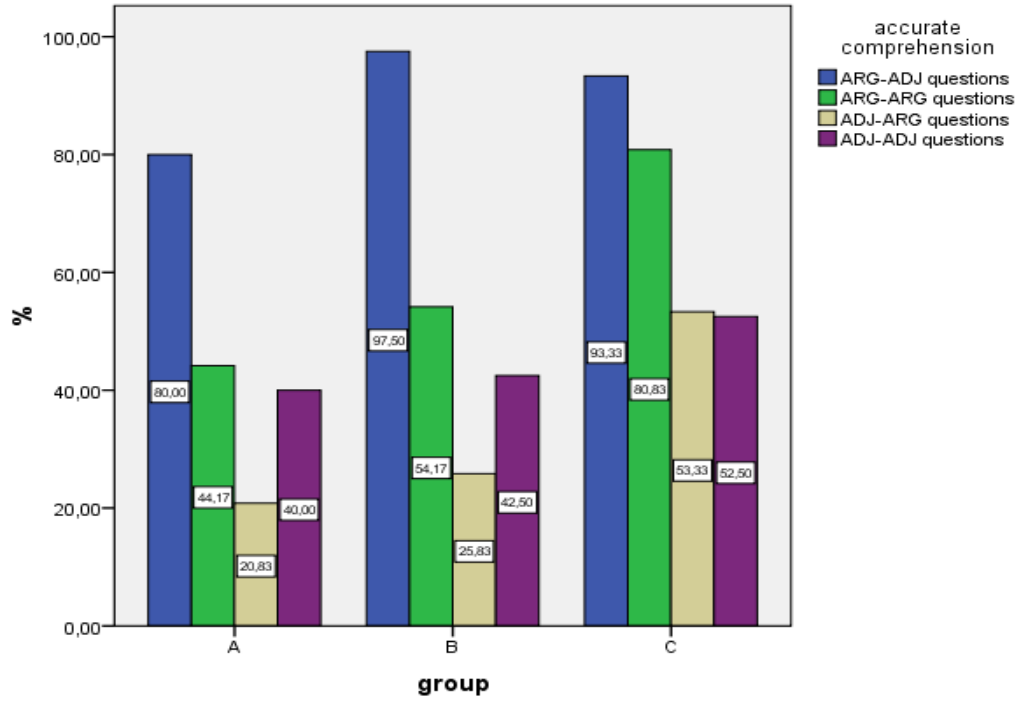


Figure 3. Accurate Performance in the Comprehension of Wh-Wh Questions



In both types of questions accurate performance increased with age. In wh-COMP questions children were very consistent, with ADJ-COMP questions being more difficult than ARG-COMP ones in the two younger groups, especially in A. As for wh-wh questions, ARG-ADJ ones proved to be the easiest for all groups, with ARG-ARG, ADJ-ADJ and ADJ-ARG questions following in decreasing order of difficulty.

Overall, the main within-group effects of accurate performance per question type [$F(5,435)=110.991$, $p=.000$] and of its interaction with group [$F(10,435)=4.958$, $p=.000$] were significant; in addition, the main between-group effect was also significant [$F(2,87)=11.303$, $p=.000$]. Pairwise, accurate comprehension of ARG-COMP questions at a within-group level was significantly higher than that of ADJ-COMP ones only in group A ($p=.001$). Between groups, no significant comparisons were found for ARG-COMP but only for ADJ-COMP, where a significant increase in accurate comprehension was found between groups A and B (A-B: $p=.005$, A-C: $p=.000$). Turning to wh-wh questions, in groups A and B all within-group comparisons were significant except for the comparison between ARG-ARG and ADJ-ADJ questions. In group C all comparisons were again significant, except for the comparisons between argument-extraction questions (ARG-ADJ vs. ARG-ARG) and between adjunct-extraction questions (ADJ-ARG vs. ADJ-ADJ). Between groups, the significant burst in accurate comprehension of argument-medial questions was attested between B and C (ARG-ARG: A-C: $p=.000$, B-C: $p=.000$; ADJ-ARG: A-C: $p=.000$, B-C: $p=.002$), whereas in ARG-ADJ questions between A and B (A-B: $p=.003$, A-C: $p=.028$). As for ADJ-ADJ ones, no significant comparisons were attested as their accurate comprehension rates remained at similar levels across groups.

On the production side, two types of questions were targeted: SD argument and adjunct questions, and LD argument and adjunct ones. The following two figures illustrate the three groups' performance in target SD and target LD question production.

As shown in figures 4 and 5, accuracy rates became increasingly higher in both argument and adjunct question production, but they were higher in SD compared to LD questions. In addition, in target SD extraction, subject questions had the higher accurate production rates for the two older groups, whereas in target LD extraction this was the case for adjunct questions across groups.

Figure 4. Accurate Performance in the Production of Target Short-Distance Questions

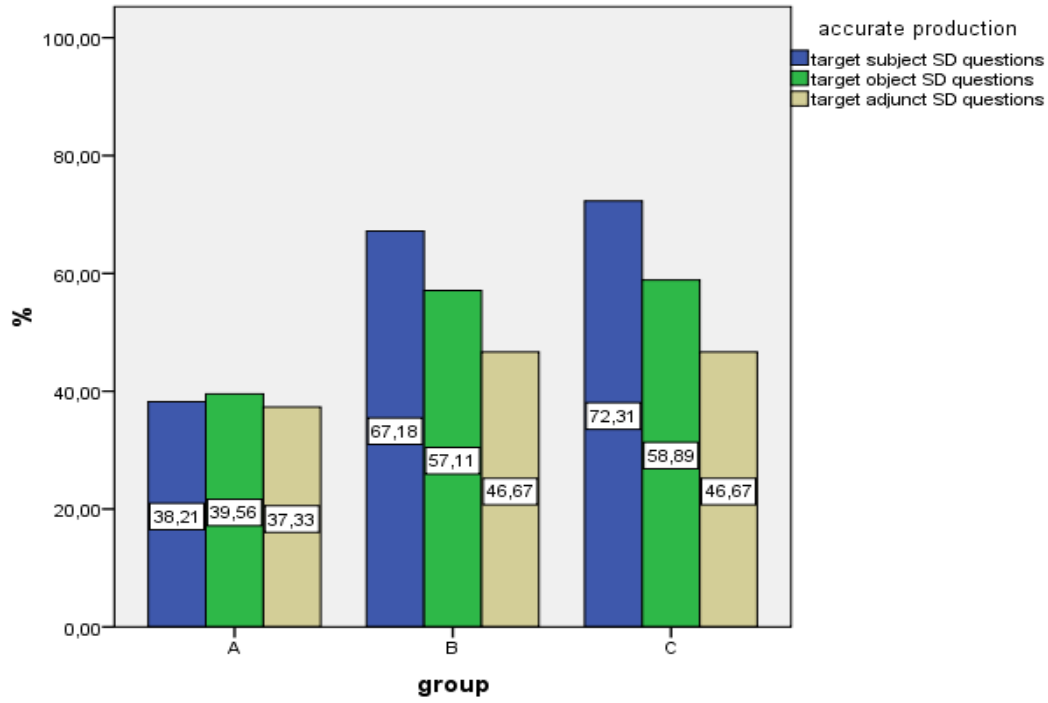
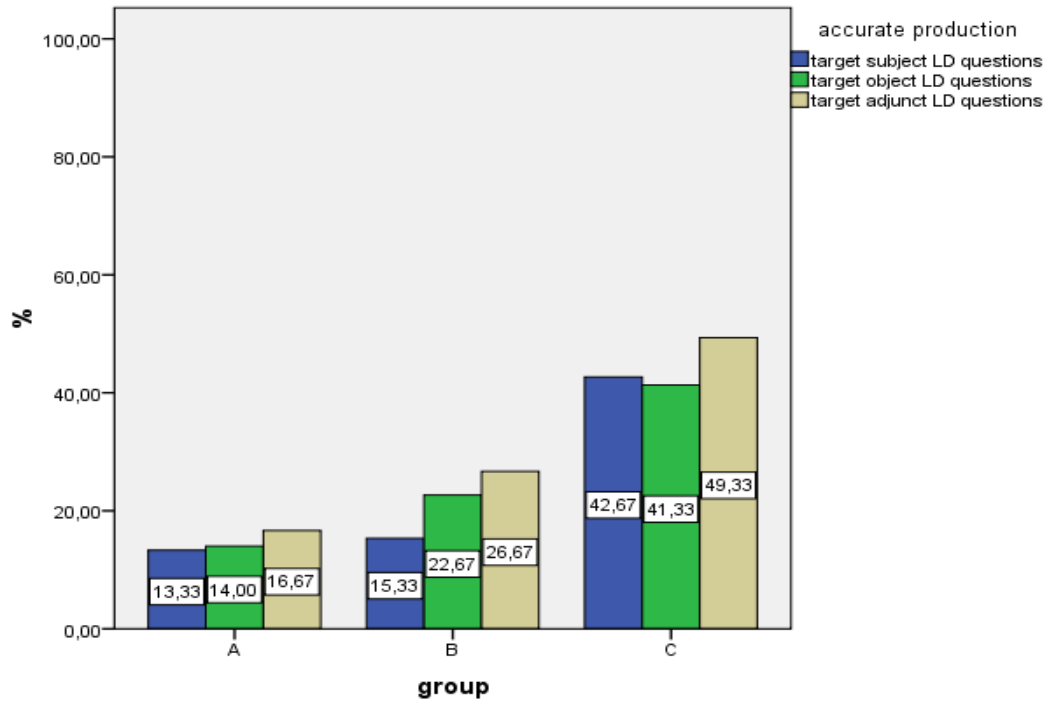


Figure 5. Accurate Performance in the Production of Target Long-Distance Questions



Within-subjects, the main effect of extraction site was significant in both target SD [F(2,174)=16.406, p=.000] and target LD [F(2,174)=7.240, p=.001] production, while the effect of extraction site by group interaction was significant only in target SD

[F(4,174)=3.838, p=.005], not in target LD [F(4,174)=1.295, p=.274]. Between subjects, however, the main effect of group was significant in both types of questions [SD: F(2,87)=8.338, p=.000, LD: F(2,87)=9.696, p=.000]. As for pairwise comparisons, within-group analyses revealed no significant differences between the various extraction sites in the youngest children's accurate SD and LD questions. For group B and C children, accurate subject, object and adjunct SD extraction rates all differed significantly from one another (B: subject-object: p=.001, subject-adjunct: p=.001, object-adjunct: p=.030; C: subject-object: p=.000, subject-adjunct: p=.000, object-adjunct: p=.015), whereas accurate LD production rates were significantly lower for subject compared to object and adjunct extraction (subject-object: p=.029; subject-adjunct: p=.004) in group B, and significantly higher for adjunct compared to subject and object extraction in group C (adjunct-subject: p=.075, subject-adjunct: p=.007). At a between-group level, accurate production with regard to all extraction sites increased significantly between A and B in target SD (subject: A-B: p=.001, A-C: p=.000; object: A-B: p=.009, A-C: p=.005; adjunct: A-B: p=.064, A-C: p=.064), and between B and C in target LD questions (subject: A-C: p=.000, B-C: p=.000; object: A-C: p=.001, B-C: p=.017; adjunct: A-C: p=.000, B-C: p=.005), which indicates clearly the greater ease of children with target SD compared to target LD question production.

As for inaccurate performance, the error types attested are broadly defined and illustrated below.

Table 1. Inaccurate Performance in Question Comprehension vs. Question Production: Error Analysis

		GROUP A		GROUP B		GROUP C	
Condition	***Error Type***	***No***	***Mean***	***No***	***Mean***	***No***	***Mean***
Question comprehension	Island violation	33/298	11.07%	18/233	7.73%	22/159	13.84%
	Medial-wh interpretation	185/298	62.08%	153/233	65.66%	73/159	45.91%
	Irrelevant interpretation	80/298	26.85%	62/233	26.61%	64/159	40.25%
Question production	Non-target landing site (SD → LD, LD → SD)	518/991	52.27%	530/754	70.29%	487/623	78.17%
	Non-target wh	35/991	3.53%	60/754	7.96%	50/623	8.02%
	Isolated wh	408/991	41.17%	147/754	19.49%	73/623	11.72%
	Irrelevant question	18/991	1.82%	10/754	1.33%	10/623	1.61%
	No response	12/991	1.21%	7/754	0.93%	3/623	0.48%

With regard to question comprehension, it is evident that medial-wh interpretations constituted the main type of error, with irrelevant interpretations and island violations following. As for question production, non-target landing site was the dominant error, followed by isolated wh production especially in group A. The error types presented here are very general categories that can be further specified into more concrete ones. In question comprehension, 'island violation' refers both to questions with a wh- and a negative island, while 'irrelevant interpretation' to questions with a wh-island and to questions without one in sum. Moreover, all error types in question production concern target SD and target LD questions together. In a more fine-grained analysis, the

distribution across particular types of questions of the errors presented in table 1, is provided in the following tables 2-5[4].

Question Comprehension

Table 2. Inaccurate Performance in the Comprehension of Wh-COMP Questions: Error Analysis

		GROUP A		GROUP B		GROUP C	
Condition	***Error Type***	***No***	***Mean***	***No***	***Mean***	***No***	***Mean***
ARG-COMP questions	Irrelevant interpretation	7/7	100%	4/4	100%	8/8	100%
ADJ-COMP questions	Island violation	13/33	39.39%	6/13	46.15%	4/8	50%
	Irrelevant interpretation	20/33	60.61%	7/13	53.85%	4/8	50%

Table 3. Inaccurate Performance in the Comprehension of Wh-Wh Questions: Error Analysis

		GROUP A		GROUP B		GROUP C	
Condition	***Error Type***	***No***	***Mean***	***No***	***Mean***	***No***	***Mean***
ARG-ADJ questions	Island violation	6/24	25%	1/3	33.33%	3/8	37.50%
	Medial-wh interpretation	11/24	45.83%	1/3	33.33%	2/8	25%
	Irrelevant interpretation	7/24	29.17%	1/3	33.33%	3/8	37.50%
ARG-ARG questions	Island violation	0/68	0%	0/57	0%	3/23	13.04%
	Medial-wh interpretation	64/67	95.52%	54/55	98.18%	19/23	82.61%
	Irrelevant interpretation	3/67	4.48%	1/55	1.82%	1/23	4.35%
ADJ-ARG questions	Island violation	2/95	2.10%	1/89	1.12%	1/56	1.79%
	Medial-wh interpretation	68/95	71.59%	56/89	62.92%	18/56	32.14%
	Irrelevant interpretation	25/95	26.31%	32/89	35.96%	37/56	66.07%
ADJ-ADJ questions	Island violation	12/72	16.67%	10/69	14.49%	11/57	19.30%
	Medial-wh interpretation	42/72	58.33%	42/69	60.87%	34/57	59.65%
	Irrelevant interpretation	18/72	25%	17/69	24.64%	12/57	21.05%

[4] The error types in tables 2-5 are presented per condition. Therefore, in order to gather, for example, the youngest children's 33 instances of island violation during question comprehension (see table 1), what needs to be done is add together all the instances of island violation that are presented per question type in tables 2 and 3.

Question Production

Table 4. Inaccurate Performance in the Production of Target Short-Distance Questions: Error Analysis

		GROUP A		GROUP B		GROUP C	
Condition	***Error Type***	***No***	***Mean***	***No***	***Mean***	***No***	***Mean***
Target: subject questions	LD	25/241	10.37%	44/128	34.38%	59/108	54.63%
	Non-target wh	18/241	7.47%	15/128	11.72%	16/108	14.81%
	Isolated wh	191/241	79.25%	65/128	50.78%	27/108	25%
	Irrelevant question	5/241	2.07%	2/128	1.56%	6/108	5.56%
	No response	2/241	0.83%	2/128	1.56%	0/108	0%
Target: object questions	LD	92/272	33.82%	116/193	60.10%	151/185	81.62%
	Non-target wh	7/272	2.57%	12/193	6.22%	9/185	4.87%
	Isolated wh	157/272	57.72%	57/193	29.53%	21/185	11.35%
	Irrelevant question	11/272	4.04%	5/193	2.59%	4/185	2.16%
	No response	5/272	1.84%	3/193	1.55%	0/185	0%
Target: adjunct questions	LD	31/94	32.98%	48/80	60%	50/80	62.50%
	Non-target wh	2/94	2.13%	4/80	5%	5/80	6.25%
	Isolated wh	59/94	62.76%	25/80	31.25%	25/80	31.25%
	Irrelevant question	2/94	2.13%	3/80	3.75%	0/80	0%

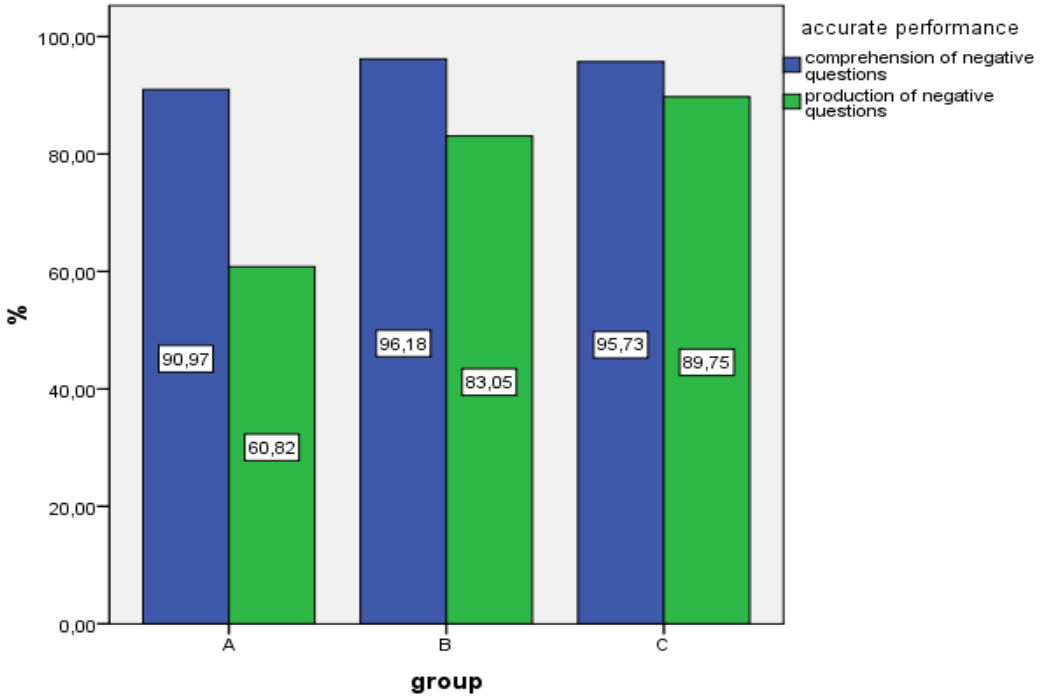
Table 5. Inaccurate Performance in the Production of Target Long-Distance Questions: Error Analysis

		GROUP A		GROUP B		GROUP C	
Condition	***Error Type***	***No***	***Mean***	***No***	***Mean***	***No***	***Mean***
Target: subject questions	SD	121/130	93.08%	112/127	88.19%	80/86	93.02%
	Non-target wh	7/130	5.38%	14/127	11.02%	5/86	5.81%
	Isolated wh	0/130	0%	0/127	0%	0/86	0%
	No response	2/130	1.54%	1/127	0.79%	1/86	1.16%
Target: object questions	SD	126/129	97.67%	110/116	94.83%	82/88	93.18%
	Non-target wh	0/129	0%	6/116	5.17%	5/88	5.68%
	Isolated wh	1/129	0.78%	0/116	0%	0/88	0%
	No response	2/129	1.55%	0/116	0%	1/88	1.14%
Target: adjunct questions	SD	123/125	98.40%	100/110	90.91%	65/76	85.52%
	Non-target wh	1/125	0.80%	9/110	8.18%	10/76	13.16%
	Isolated wh	0/125	0%	0/110	0%	0/76	0%
	No response	1/125	0.80%	1/110	0.91%	1/76	1.32%

On the whole, question comprehension was found to develop earlier than question production, with significant differences attested in all groups. Therefore, the main prediction is, in general terms, borne out. Before ending this section, however, it is interesting to see whether this pattern remained the same when dealing with negative questions exclusively.

Figure 6 below presents the average accurate performance rates of all groups with respect to the comprehension of negative questions and the production of negation in target negative contexts, irrespective of accurate SD/overuse of LD extraction counts[5].

Figure 6. Accurate Performance in Negative Questions



As depicted, accurate performance in negative questions was the dominant pattern from the youngest age, with accurate performance rates increasing with age. Within groups it is evident that children, especially in the two younger groups, performed much better in the comprehension rather than in the production of negative questions.

The main effect of negation within subjects was significant [F(1,87)=29.992, p=.000], and so was the effect of negation by group interaction [F(2,87)=4.729, p=.011]. As for the main between-subjects effect of group, it was also found to be significant [F(2,87)=7.760, p=.001]. Pairwise, within-group comparisons revealed that accurate performance in the presence of negation was significantly better at the comprehension than the production level in groups A (p=.000) and B (p< .01), but not in C (p=.249). As for the between-group analysis, in comprehension no significant comparisons were attested as accurate performance rates were very high from the youngest test age. Turning to production, the presence of negation was not avoided in target negative questions; on the contrary, accuracy rates were above 50% from the youngest age, with significant differences attested between groups A and B (p=.002), and between groups A and C (p=.000). This suggests that a significant burst in the production of negation occurred between A and B, and the accuracy rates increased gradually thereafter.

[5] On the comprehension side, what were of interest in the present measurement were children's responses with regard to the fronted wh-element. On this ground, medial-wh and irrelevant interpretations were excluded from the measurement. On the production side, instances of isolated wh and irrelevant questions were excluded.

As a final note, it is interesting to mention that the negative questions produced by the Greek children in the present study were grammatical in their overwhelming majority. Instances of negation doubling were not attested, while instances of verb doubling were scarce and amounted to only six across groups: five in group A and one in group C. As a matter of fact, all these six instances involved doubling not only of the verb but also of the wh-element, and are listed just below.

(1) 'Pu tha pai pu dhe tha pai kanis?'
where will go-3SG where not will go-3SG no one
target: Pjos na min pai puthena simera?
(= 'Who should not go anywhere today?')

(2) 'Ti na pari ti mi pari mesa sto spiti?'
what to take-3SG what not take-3SG inside the house
target: Ti na min pari mesa sto spiti?
(= 'What should he not take in the house?)

(3) 'Pjo na pai pjo na min pai makria?'
which to go-3SG which to not go-3SG away
target: Pjo pedhi na mi fiji makria? (= 'Which child should not go away?')

(4) 'Ti na min fai o Kostas ti na fai?'
what to not eat-3SG the Kostas what to eat-3SG
target: Ti na min fai o Kostas? (= 'What should Kostas not eat?')

(5) 'Pjo na bi pjo na mi bi mesa?'
which to enter-3SG which to not enter-3SG inside
target: Pjo pedhi na mi bi mes sto spiti?
(= 'Which child should not go in the house?')

(6) 'Astruli pjo dhokimazi pjo dhe dokimazi o filos su?'
Astruli which try-3SG which not try-3SG the friend your
target: Pjo fajito dhen troi o Jack? (= 'Which food doesn't Jack eat?')

The rest of the negative questions produced across groups were grammatical, with two of them actually instantiating the use of a cleft construction[6].

(7) 'Pjos ine aftos pu dhe bori na pai?'
who be-3SG he that not can-3SG to go-3SG
'Who is the one that cannot go?'
target: Pjos dhe tha voithisi ti mama? (= 'Who will not help mum?')

[6] As a matter of fact, instances of cleft questions were only three throughout the production data. The two of them are the ones listed in (7) and (8), while the third instance of wh-cleft was produced by one of the oldest children in a target affirmative context and is presented below.

'Pjos itan pu xoreve orea sti jiortula?'
who was-3SG that danced-3SG well at the party
'Who was the one who danced well at the party?'
target: Pjos xorepse orea sti jiorti? (= 'Who danced well at the party?')

(8) ‘Pjo itan to zoaki pu de xorepse para poli kala?’
which was-3SG the animal that not danced-3SG very very well
‘Which was the animal that did not dance very well?’
target: Pjo zoo dhen xorepse katholu kala?
(= ‘Which animal did not dance well?’)

On the whole concerning negative questions, the above results indicate that negation effects were present in child grammar from the youngest test age, with comprehension being better compared to production. In other words, the more general pattern of higher accuracy rates in comprehension than production seems to be confirmed even when looking to negative questions in isolation.

4. Discussion and Conclusions

To sum up with regard to children’s performance on the acquisition of wh-questions, the main result was that the comprehension of wh-questions showed overall significantly higher accurate performance rates than the production of wh-questions from the youngest test age. Besides, it is important to note that children’s better performance in comprehension over production is also reflected at a more fine-grained level. That is, accurate performance in comprehension of all wh-COMP and wh-wh questions, with the exception of ADJ-wh ones, either was high/top[7] by the youngest test age or became high/top at a certain age. On the production side, accurate performance displayed some degree of development as well: it increased significantly by group B in target SD and by group C in target LD questions. However, it never reached a level of high/top performance, thus being in contrast with accurate performance in comprehension.

Regarding accurate performance per extraction site, the following tendencies were observed. As far as comprehension is concerned, accurate performance in wh-COMP questions was high/top in the two older groups but significantly lower in adjunct (average performance) compared to argument extraction (top performance) for the youngest children. In the presence of a wh-island (i.e. wh-wh questions), however, accurate comprehension rates were generally lower. In the two younger groups, ARG-ADJ were the significantly easiest (high/top performance) and ADJ-ARG the significantly most difficult questions to comprehend (low performance), with ARG-ARG and ADJ-ADJ falling in-between with decreasingly low-level rates of accurate performance. As for the oldest group, all argument questions (ARG-ADJ, ARG-ARG) were significantly easier (high/top performance) than all adjunct ones (ADJ-ARG, ADJ-ADJ) (low performance). On the whole, then, it seems to be the case that the comprehension of argument-extraction questions caused less problems than the comprehension of adjunct-extraction ones to the 4-to-7 year old Greek children that participated in this study. At the production level, where target LD questions were generally more difficult to produce than target SD questions across all test ages, the argument/adjunct extraction site did not affect the youngest children’s accurate performance, which remained at low levels in all conditions. Yet, in groups B and C a

[7] Such characterizations of performance throughout this section refer to the following specifications:
top performance: 90% <
high performance: 80% - 89.99%
average performance: 60% - 79.99%
low performance: < 59.99%

different pattern was observed: in target SD questions adjunct extraction was found to be significantly more difficult than argument extraction, whereas in target LD questions adjunct extraction was easier to produce than argument extraction.

On the basis of all these observations regarding accurate performance per extraction site, one important similarity seems to emerge between comprehension and production. Specifically, the comprehension of wh-wh questions, where grammatical interpretation of the raised wh essentially involves SD movement (cf. Thornton & Crain 1994: 228), patterns together with the production of target SD questions. That is, at both the comprehension and the production level, SD extraction of adjuncts was more problematic than SD extraction of arguments, especially in the older groups. At first sight, the increased problems with adjunct extraction may be considered to be the result of conceptual difficulties associated with the more abstract concepts expressed by adjuncts compared to arguments (e.g. Bloom et al. 1982: 1084, Thornton & Crain 1994: 248-249). Interestingly, however, a reverse pattern applied to LD extraction at the production level, where adjunct questions were less problematic than argument questions. From a closer look at a deeper level, this discrepancy pattern between arguments and adjuncts may be argued to reflect an attempt on the part of the children to save on processing resources. Given that arguments are projected by the lexicon and need to be attached into the parse tree under the dictation of grammatical principles while adjuncts are only optionally attached to the verb by a more global principle (e.g. Full Interpretation) (Pritchett 1991: 327), it follows that the formation of wh-argument chains is more demanding in terms of processing than the formation of wh-adjunct chains. Modality of performance (comprehension/production) and distance of extraction (short/long) are then considered to play a counterbalancing role, with accurate argument chains being more prominent in target SD production and question comprehension, and accurate adjunct chains more prominent in target LD question production[8].

To summarize so far, these findings suggest that children's performance follows consistently resource-saving strategies. The significantly higher accurate performance in question comprehension than in question production as well as the attainment of high/top performance in the former as opposed to the latter, are well explained under the assumption that comprehension necessitates the employment of fewer processing resources than production. Furthermore, even at an extraction-specific level, the significantly higher accurate performance on argument and adjunct extraction in comprehension/SD production and in LD production respectively, may also be treated as evidence in support of children's employment of resource-saving strategies.

In addition, it is interesting to note, in passing, that in target SD and target LD production, accurate performance rates in argument extraction did not differ significantly between subject and object questions in the youngest group. Yet, significant differences were attested in the two older groups: subject extraction was significantly easier in target SD questions for groups B and C, whereas object extraction was significantly easier in target LD questions, but only for group B[9]. Our findings from the youngest children contradict relevant crosslinguistic ones (e.g. Fahn

[8] Remember that comprehension is less costly in terms of processing than production, and so is SD over LD production.
[9] For group C children, no significant differences were observed between subject and object extraction.

2003, Guasti et al. 2011), which have been explained in terms of length-of-wh-chain effects on children's parser: subject chains can be formed earlier than object chains, since the latter always include an intervener (typically the subject) between the filler and the gap (Friedmann et al. 2009: 81). On this ground, subject chains require less memory load in order to get processed, and hence they are preferred over object chains (e.g. Gibson et al. 1994). According to Guasti et al. (2011), this subject/object asymmetry disappears in Greek because the morphological case marking on the wh-expressions nullifies the presence of any intervention effect. This seems to hold for the youngest children in the present study, but not for the older ones. For the older children, length-of-wh-chain effects similar to those attested crosslinguistically seem to be in operation in target SD questions and thus to override case-marking effects; as for the greater ease with object extraction in target LD questions, this might be explained on pragmatic grounds: object questions are less marked than subject ones in terms of focus, and this reduced pragmatic markedness may counterbalance the greater processing cost associated with LD dependencies, thus leading to higher accuracy in the production of target object compared to target subject LD questions (cf. Stromswold 1995).

Turning to negative questions as a final point, the early Greek data showed that the presence of negation was generally not avoided in target negative questions, since accurate performance rates had become high by group B. It is worth underlining that almost all negative questions produced by the Greek children were grammatical. At a crosslinguistic level, early English data has shown that children's negative questions often involve some type of doubling, like doubling of the auxiliary or doubling of the auxiliary along with negation (Guasti et al. 1995, Hiramatsu 2003, Thornton 1995). On the contrary, early Italian negative interrogatives are adult in form, with doubling occurrences never being attested (Guasti 1996). Given these two trends in the production of negative questions, early Greek seems to pattern together with Italian, since the only observed instances of doubling included wh and verb doubling and amounted to only six across groups. Drawing on the claim that children initially assume that negation must stay in a V-related projection (cf. Guasti 1996, Guasti et al. 1995), it is argued here that Greek children, on a par with their Italian but unlike their English peers, seem to hypothesize correctly from start that the Neg-Criterion must be satisfied inside CP and not IP. Unlike English, all verbs can raise to C in Greek and Italian, and hence CP is the V-related projection where the negation feature is checked. For this reason, Greek children locate negation inside CP from start and produce adult-like negative interrogatives[10]. The highly grammatical negative questions provide validation of the claim on children's initial assumption to place negation in a V-related projection, which attests, in turn, to that children adopt the most restrictive hypotheses possible at each stage of processing (cf. Guasti et al. 1995).

On the whole, all these findings converge to a common point: in child grammar, the most economical options prevail, as dictated by the natural economy principle that

[10] Contrary to Greek children, Italian children have to *raise* negation to CP in order to produce grammatical negative questions. This difference between Greek and Italian stems from the fact that negation occupies a higher structural position in the former compared to the latter language. That is, in Greek NegP is located between the two lower C heads C_{OP} and C_M within a split-CP domain (see Roussou 2000: 79 for more details), while in Italian negation is located within IP (Guasti et al. 1995, Guasti 1996). As a result of this difference, adult-like negative interrogatives in Greek, unlike Italian, do not involve raising to but merely retaining of negation inside CP.

permeates the operations of the language acquisition device. In other words, an economy-based hierarchy of choices is in play, with the least marked strategies being preferred over the most marked ones. What is essential to underline here is that this economy-based hierarchy of choices is not triggered exclusively on syntactic grounds. As suggested in the above discussion, semantic and pragmatic factors also seem to affect the formation of this hierarchy. Thus Jakubowicz's (2005) Derivational Complexity Hypothesis, according to which less complex derivations are correctly spelled out at the PF interface before more complex ones during language development, can be seriously challenged. Children's processing ability may lag behind not at the level of form but at the level of meaning integration. In other words, maturation in terms of processing, which constitutes an extra-linguistic domain, is involved, and this maturation concerns semantics rather than syntax. On the whole, then, maturation of a non-linguistic ability is hypothesized, which is in line with a continuity account of language acquisition (e.g. Hyams 1986, 1994, Pinker 1984).

References

Abdulkarim, L., T. Roeper and J. de Villiers, 1997: "Negative Islands in Language Acquisition", paper presented at *New Perspectives on Language Acquisition: Minimalism and Pragmatics*, University of Massachusetts, Amherst, June 2 - 4, 1997.

Bloom, L., S. Merkin and J. Wooten, 1982: ""Wh"-Questions: Linguistic Factors That Contribute to the Sequence of Acquisition", *Child Development* 53(4), pp. 1084-1092.

Cheng, L. L.-S., 1991: *On the Typology of Wh-Questions*. PhD Dissertation, MIT.

Cheng, L. L.-S., 1997: *On the Typology of Wh-Questions*. New York, London: Garland Publishings.

Crain, S. and R. Thornton, 1998: *Investigations in Universal Grammar. A Guide to Experiments on the Acquisition of Syntax and Semantics*. Cambridge, Massachusetts: MIT Press.

Fahn, R.-L. S., 2003: "Chinese-Speaking Children's Productions of *Wh*-Questions", *Concentric: Studies in English Literature and Linguistics* 29(2), pp. 82-117.

Fanselow, G., R. Kliegl and M. Schlesewsky, 1999: "Processing Difficulty and Principles of Grammar", in S. Kemper and R. Kliegl (eds.), *Constraints on Language: Aging, Grammar and Memory*, pp. 171-201. Dordrecht: Kluwer.

Frazier, L. and G. B. Flores d' Arcais, 1989: "Filler Driven Parsing: A Study of Gap Filling in Dutch", *Journal of Memory and Language* 28, pp. 331-344.

Friedmann, N., A. Belletti and L. Rizzi, 2009: "Relativized relatives: Types of intervention in the acquisition of A-bar dependencies", *Lingua* 119, pp. 67-88.

Gibson, E., G. Hickok and C. T. Schütze, 1994: "Processing Empty Categories: A Parallel Approach", *Journal of Psycholinguistic Research* 23(5), pp. 381-405.

Guasti, M. T., 1996: "Acquisition of Italian interrogatives", in H. Clahsen (ed.), *Generative perspectives on language acquisition*, pp. 241-269. Amsterdam, Philadelphia: John Benjamins.

Guasti, M. T., 2000: "An excursion into interrogatives in Early English and Italian", in M.-A. Friedemann and L. Rizzi (eds.), *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics*, pp. 105–128. London, New York: Longman.

Guasti, M. T., R. Thornton and K. Wexler, 1995: "Negation in children's questions: The case of English", *Proceedings of the 19th Annual Boston University Conference on Language Development*, pp. 228-239.

Guasti, M. T., C. Branchini and F. Arosio, 2011: "Interference in the production of Italian subject and object wh-questions", *Applied Psycholinguistics*.

Hiramatsu, K., 2003: "Children's Judgments of Negative Questions", *Language Acquisition* 11(2), pp. 99-126.

Hyams, N., 1986: *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Hyams, N., 1994: "V2, Null Arguments and COMP Projections", in T. Hoekstra and B. D. Schwartz (eds.), *Language Acquisition Studies in Generative Grammar, pp. 21-55*. Amsterdam: John Benjamins.

Jakubowicz, C., 2005: "The language faculty: (Ab)Normal development and interface constraints", paper presented at *GALA 2005*, Siena.

Philip, W. and J. de Villiers, 1992: "Monotonicity and the acquisition of weak wh-islands", in E. V. Clark (ed.), *Proceedings of the 24th Annual Child Language Research Forum*, pp. 99-111. Stanford, California: CSLI Publications.

Pinker, S., 1984: *Language Learnability and Language Development*. Cambridge, Massachusetts: Harvard University Press.

Pritchett, B. L., 1991: "Subjacency in a Principle-Based Parser", in R. C. Berwick et al. (eds.), *Principle-Based Parsing: Computation and Psycholinguistics*, pp. 301-345. Kluwer Academic Publishers.

Roeper, T. & J. de Villiers, 1992: "Ordered Decisions in the Acquisition of Wh-questions". In J. Weissenborn, H. Goodluck & T. Roeper (eds.), *Theoretical Issues in Language Acquisition: Continuity and Change in Development*, pp. 191-236. Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Stromswold, K., 1995: "The Acquisition of Subject and Object Wh-Questions", *Language Acquisition* 4(1), pp. 5-48.

Theophanopoulou-Kontou, D., 1991: "Nisida-erot (wh-island) ke metakinisi: Prokatarktikes paratiriseis sta NE") ("Wh-island and movement: Preliminary remarks in MG"), *Studies in Greek Linguistics* 12, pp. 195-215. Thessaloniki: AUTH.

Thornton, R., 1990: *Adventures in long-distance moving: the acquisition of complex wh-questions*. PhD Dissertation, University of Connecticut.

Thornton, R., 1995: "Referentiality and Wh-Movement in Child English: Juvenile D-Linkuency", *Language Acquisition* 4(1), pp. 139-175.

Thornton, R., 1996: "Elicited Production", in D. McDaniel, C. McKee and H. Smith Cairns (eds.), *Methods for Assessing Children's Syntax*, pp. 77-102. Cambridge, Massachusetts: MIT Press.

Thornton, R. and S. Crain, 1994: "Successful cyclic movement", in T. Hoekstra and B. Schwartz (eds.), *Language Acquisition Studies in Generative Grammar*, pp. 215-252. Amsterdam, Philadelphia: John Benjamins.

de Villiers, J., T. Roeper and A. Vainikka, 1990: "The acquisition of long-distance rules", in L. Frazier and J. de Villiers (eds.), *Language Processing and Language Acquisition*, pp. 257-297. Dordrecht: Kluwer.