

The ‘Phased’ Learnability of Long Wh-questions

Jacqueline van Kampen

UiL OTS, Utrecht University

1. Long wh-questions in acquisition

This paper discusses the learnability of long wh-questions. Taking longitudinal data from child Dutch as a case study, I will show that the acquisition of long wh-movement has been thoroughly prepared in previous acquisition steps. Each step defines a local relation, and thereby a local domain, that is preserved in the next acquisition step. The long wh-questions in child language appear first with an intermediate pronoun in C° (Thornton 1990, Van Kampen 1997). The Dutch data show how the intermediate pronoun relies on the (previously acquired) relative paradigm. This is so obvious, since the Dutch relatives offer an apparently irregular mixture of *d*-pronouns and *w*-pronouns. The present view on the learnability of A-bar chains will lead to the following conclusions: (i) Long wh-movement is successive cyclic from the very beginning on; (ii) Pied-piping of wh-phrases follows from the movement of a <+wh> head plus the preservation of its licensing contexts only; (iii) Syntactic islands as such are not “learned”. They follow automatically from a non-overlap of local movement domains.

Child language is the area where limited data already reveal the general principles of grammar. For example, data from the acquisition of A-bar chains (Van Kampen 1997, 2004) may support the claims about locality in (1). For related questions about locality and acquisition, see De Villiers et al. (2007).

- (1) Grammar is acquired from the most local patterns possible and the final result preserves much of that locality.
- All movement, e.g. wh-movement, is learned in a short step first. Long-distance movements follow from short steps and a fortuitous overlap of initial localities.
 - Islands need not be learned. They follow from the fortuitous non-overlap of such localities.

The type of data that will support the claims in (1) are given in (2). The sentences in (2) are long wh-questions in child Dutch with an intermediate pronoun, a *w*-pronoun in (2)a and a *d*-pronoun in (2)b.

- (2)
- welke jongen denk je *wie*<+animate> daar loopt?
which boy think you who there walks?
(which boy do you think is walking over there?)
 - welke jongen denk je *die*<+gender> ik leuk vind?
which boy think you that I nice find?
(which boy do you think I like?)

Long wh-questions with an intermediate pronoun appear in the acquisition of various languages (Thornton 1990 for English, Van Kampen 1997 for Dutch, Strik 2006 for French, Gutiérrez 2006 for Basque and Spanish). Although the type with intermediate pronoun is attested in the adult speech for a number of languages (Van Kampen 1997 and references therein), it does not do so in the (standard) adult speech of any of the languages just mentioned. Just for that reason, it is remarkable in (2) that child Dutch varies the intermediate pronoun as *die/wie*, whereas adult standard Dutch has no intermediate pronoun and only fits in the neutral complementizer *dat* (‘that’). Let me add that, in the corpora considered, this spell-out of the intermediate C°, although attested in Dutch dialects (Barbiers, Koenenman & Lekakou 2007), was non-existent in the maternal input, but absolutely dominant and quite long-lasting in the language of the child. Hence, we have here the paradoxical fact that the child introduces spontaneously a variation not present in the input.

* The research for this paper was supported by NWO grant 360-70-290. An extensive version of this paper appeared as Van Kampen (2008).

An important circumstance regarding the intermediate pronouns in (2) is that the acquisition of long wh-questions in child Dutch follows the earlier acquisition of relative pronouns. Relative pronouns show an agreement pattern with an antecedent. The relative paradigm in Dutch is quite irregular. It is an intricate mixture of *d*-pronouns and *w*-pronouns (Van Kampen 2007). Significantly, that system has been firmly acquired a year before the chains with intermediate pronouns appear. The evidence that it is the same system that gets active is as striking as the relative paradigm is irregular in Dutch, because these irregularities reappear on the intermediate pronoun paradigm. This suggests that the intermediate pronoun is to be analyzed as an “A-bar pronoun” spelled out under local agreement with the moved wh-phrase (Van Kampen 1997: chap.6, cf. Thornton 1990, Crain and Thornton 1994). The variation of the intermediate pronoun in child Dutch (*wat* ‘what’, *wie* ‘who’, *waar* ‘where’, *welk* ‘which’, *die* ‘that’, *dat* ‘that’) follows the relative agreement paradigm and can be explained within that perspective.

Section 2 will exemplify the irregular variation in the relative paradigm. Section 3 settles certain syntactic issues, ‘left edge accessibility’, the acquisition of Ross’ (1967) Left Branch Condition, and the simultaneous acquisition of obligatory pied-piping. Section 4 offers the empirical core of the paper. The intermediate A-bar pronouns in C⁰ can be derived by the already acquired rule for relative agreement. This leads to an obvious point and a curious prediction. The obvious point is that the intermediate C⁰ content follows from antecedent agreement. The curious prediction runs this way. The irregularities of relative agreement should reappear on the intermediate pronouns. This prediction is correct and far from trivial. The support for a multiple short step analysis of long wh-movement could not be better, since the options that the relative paradigm allows are open at each intermediate point, but they are not present in the adult input. In order to acquire the adult system, the child only needs to suppress the spell-out of the agreement and use a complementizer instead. In this view, the top of the chain must remain the most specified element, since it is the antecedent in all intermediate moves. Finally, section 5 offers the learnability perspective. Locality appears an inevitable ingredient for the learnability of long-distance relations. The necessity of ‘learning from local steps’ is supported by the acquisition data. Since the child’s acquisition steps show the locality in overt syntax, successive cyclic movement is the best hypothesis for the observed data.

2. Morphological preliminaries: The relative pronoun paradigm

Pronouns are referential indicators <+D> (Postal 1966). A-bar pronouns are pronouns that must end up in a sentence-initial CP position (Van Kampen 1997: 92ff). They represent pronominal categories with the additional feature <+C>. The best examples of inherently A-bar pronouns are *w*-pronouns in root questions and relative pronouns. The V2nd languages have in addition an A-bar pronoun used as a topic in root sentences, the *d*-pronoun, see (3).

- (3) ze zag daar een jongen. *Die* (= de jongen) vond ze wel leuk
 she saw there a boy. That (= the boy) found she rather nice
 (she saw a boy over there. She thought he was rather nice)

The entire paradigm of <+C> question pronouns in Dutch is <+wh> (*wat*, *wie*, *waar*) and that of <+C> topic pronouns is <-wh> (*dat*, *die*, *daar*). See the list in (4).

- (4) A-bar pronouns in Dutch root sentences

	referent	structural case	oblique case
<i>d</i> -set	<+gender>	<i>die</i>	---
	<-gender>	<i>dat</i>	([<i>daar</i>] ... (P))
<i>w</i> -set	<+animate>	<i>wie</i>	[P <i>wie</i>]
	<-animate>	<i>wat</i>	[<i>waar</i>] ... (P)

Agreement is controlled by the <±gender> and <±animate> properties of the referent. The topic *d*-pronouns have an antecedent and they are sensitive to phi-features <±gender> of that antecedent. In Dutch, singular nouns can be <+gender> or <-gender>. This feature determines the choice of the definite article, either *de vrouw* <+gender> (‘the woman’) or *het huis* <-gender> (‘the house’). The plural definite nouns are always *de* (*de vrouwen*, *de huizen*). Slightly different from the traditional

descriptions, I consider “number” as adding the feature <+gender>/de. The oblique pronoun *daar* is not sensitive to gender. The question *w*-pronouns have no syntactic antecedent, but they indicate nevertheless whether the answer must be <±animate>.

A-bar pronouns are also related to an argument position. The A-bar pronouns in Dutch express structural versus oblique properties. This <±oblique> feature is not related to the antecedent, but to the A(argument)-position, see the examples in (5). Note that *kast* (‘cupboard’) is a <+gender> *de*-noun.

- (5) a. [op welke kast]<+oblique> *die*<+gender> jij hebt t_d gekocht ligt al dat stof?
 on which cupboard that you have t_d bought lies all that dust?
 b. *de kast*<+gender> *waar*<+oblique> al dat stof op t_{wh} ligt heb jij gekocht
 the cupboard where all that dust on t_{wh} lies have you bought

The general properties of A-bar pronouns are now given in (6).

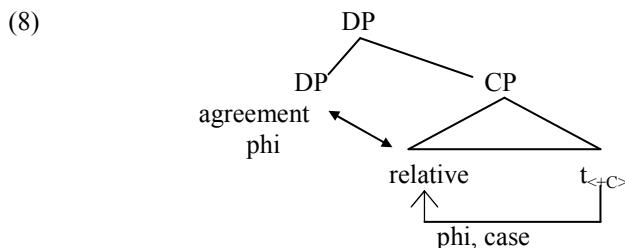
(6) *Properties of A-bar pronouns*

- are characterized by <+D>, <+C> and <±wh>.
- express phi-features for <±gender> (*d*-pronouns) and <±animate> (*w*-pronouns) of the antecedent/referent.
- express the <±oblique case> of the argument position.

Relative pronouns are A-bar pronouns as well. They have the characteristics in (7).

- (7) a. Their position is a sentence-initial <+C> position.
 b. Their form is partly taken from the root *w*-pronouns <+C, +wh>, and partly from the root *d*-pronouns <+C, -wh> in languages that have them, like V2nd Dutch.

The *d*-option for relative pronouns is characteristics of V2nd languages (Van Kampen 2007). Relative pronouns are A-bar pronouns that agree in phi-features with the antecedent. The relative pronoun is an A-bar pronoun that relates to the case properties (<±oblique>) of the argument gap t_{<+C>} and to the phi-features of the DP that is the sister of its CP-projection, see (8).



The pronominal paradigm of the relative in Dutch is a mixture of *d*-pronouns and *w*-pronouns. If it is possible to express gender agreement with the antecedent, the *d*-pronoun is selected as in (9)a,b. Since the oblique *daar* is not sensitive to gender, *daar* is not selected as a relative, see (9)c.

- (9) Dutch relative pronouns with <±gender> agreement. The *d*-set comes in: *die*, *dat*, **daar*

	structural case	oblique case
<+gender>	<i>die</i>	---
<-gender>	<i>dat</i>	*[<i>daar</i>] ... (P)

- de jongen die*<+gender> ik leuk vind
(the boy that I like)
- het huis/het meisje dat*<-gender> ik leuk vind
(the house that I like)
- *in het huis *daar*<+oblique> ik woon
(in the house where I live)

If it is not possible to express gender agreement, the *w*-pronoun, sensitive to $\langle \pm \text{animate} \rangle$, is selected as in (10). This includes ‘fused’ relatives when there is no antecedent, (10)a,b. In contrast to the question *w*-pronoun, cf. (4), the oblique pronoun *waar* can be used with $\langle \pm \text{animate} \rangle$ antecedents, (10)c,d. Parallel to the question *w*-pronoun, see (5), only pronouns that are marked for $\langle + \text{animate} \rangle$ (i.e. *wie*) can realize [P pronoun] (10)e.

- (10) Dutch relative pronouns with $\langle \pm \text{gender} \rangle$ agreement blocked due to the absence of the antecedent or due to $\langle + \text{oblique} \rangle$. The *w*-set takes over: *wie*, *wat*, *waar*

	structural case	oblique case
$\langle + \text{animate} \rangle$	<i>wie</i>	[<i>waar</i>]...P [P <i>wie</i>]
$\langle - \text{animate} \rangle$	<i>wat</i>	[<i>waar</i>] ... (P)

- wie* $\langle + \text{animate} \rangle$ ik leuk vind, is het hoofd van de school
(who I like, is the head of the school)
- wat* $\langle - \text{animate} \rangle$ ik leuk vind is die bank
(what I like, is that couch)
- in het huis *waar* $\langle + \text{oblique} \rangle$ ik woon
(in the house where I live)
- het huis/de jongen *waar* $\langle + \text{oblique} \rangle$ ik verliefd *op* ben
(the house/the boy with whom I am in love)
- de jongen *op wie* $\langle + \text{animate} \rangle$ ik verliefd ben
(the boy with whom I am in love)

There are two major exceptions to the generalizations of the selection in (9)-(10). First, (11)a shows that the pronoun *die* may be used as $\langle + \text{animate} \rangle$ in relative agreement, although gender agreement (*dat*, cf. (9)b) might have been possible. This parallels with the tendency in spoken Dutch to extend *die* to $\langle - \text{gender} \rangle$ antecedents that are semantically specified for $\langle + \text{animate} \rangle$, see (11)b. In that case, semantic animacy overrules grammatical gender, which is not perceived by the speaker.

- (11) a. het meisje $\langle - \text{gender} \rangle / \langle + \text{animate} \rangle$ *die* $\langle + \text{animate} \rangle$ ik leuk vind
the girl that I nice find
(the girl that I like)
- b. neem nou het buurmeisje $\langle - \text{gender} \rangle / \langle + \text{animate} \rangle$. *Die* $\langle + \text{animate} \rangle$ vind ik leuk
take now the neighbors-girl. That find I nice
(take the girl of the neighbors. I like her)

Second, (12) shows that the pronoun *wat* may be used as with $\langle - \text{gender} \rangle$ antecedents in relative agreement. This selection of *wat* is preferred in spoken Dutch over the selection of *dat* in (9)b.

- (12) het huis/het meisje *wat* $\langle - \text{gender} \rangle$ ik leuk vind
the house/the girl what I like

Both irregularities are analyzed and explained in Van Kampen (2007) as the outcome of a selection problem related to the order of acquisition steps. The acquisition of gender is too slow to suppress the $\langle + \text{animate} \rangle$ agreements of the *w*-system.

The scheme in (13) summarizes the variation in the relative paradigm. The set in (13) covers all observations as made in the standard grammar ANS (Algemene Nederlandse Spraakkunst 1997). The acquisitional explanation of the irregularity in the relative paradigm is from Van Kampen (2007).

(13)

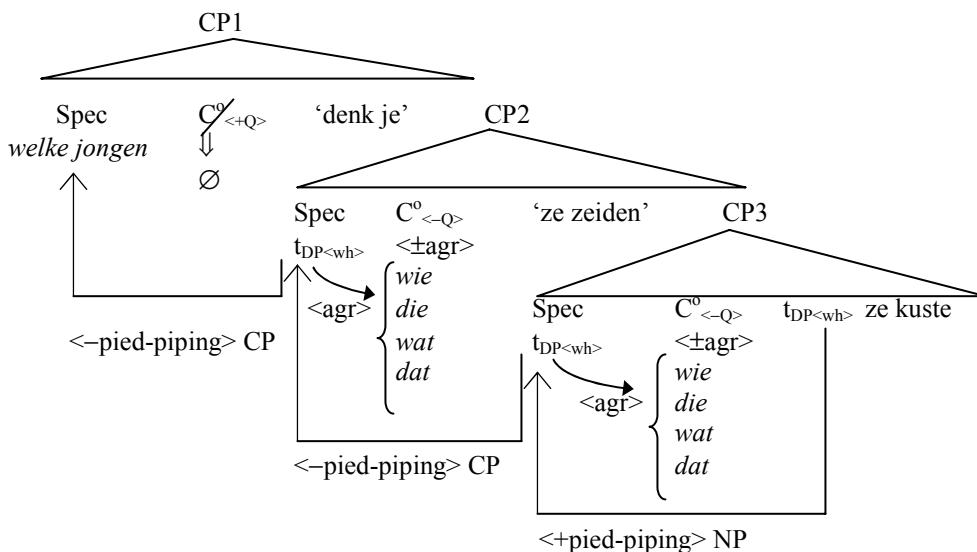
	<+gender>	<-gender>		no gender	
	gender overrules animacy	<+animate>	<-animate>	<+animate>	<-animate>
<-oblique>	de jongen (the boy) <i>die / *wie</i>	het meisje (the girl) <i>dat / wat</i> <i>die / *wie</i>	het paard (the horse) <i>dat / wat</i>		
<+oblique>				de jongen (the boy) <i>waar ... P</i> <i>P wie</i> <i>*P die</i>	het huis (the house) <i>waar ... (P)</i> <i>*daar ... (P)</i>
No antecedent				<i>wie</i>	<i>wat</i> <i>waar</i>

The factual irregularity of the A-bar relative paradigm is what matters here. It will be used as an argument to show in section 4 that the relative paradigm reappears as a filter on the A-bar agreement for the intermediate C^0 pronouns in successive cyclic long wh-chains. Below, I will first argue that the child first reconstructs the long wh-movement from the most local movement possible.

3. Syntactic preliminaries

The analysis of the intermediate A-bar pronouns as resulting from a local Spec-head agreement implies that long wh-movement is successive cyclic. It does not explain, though, why the agreement does not appear when the wh-antecedent makes landing in the final Spec.C. Nor has it been explained why, in its first move, the $D_{<+wh>}$ obeys the Left Branch Condition and pied-pipes the NP, whereas it does not pied-pipe the CP. Which mechanism switches the LBC on for NP and out for CP? See (14).

(14)



which boy think you $C^0<+agr>$ they said $C^0<+agr>$ she kissed
(which boy do you think they said she kissed)

The first move of the wh-element *welke* ('which') in (14) pied-pipes (moves along) the NP *jongen* ('boy'). This initial pied-piping continues to be relevant for the subsequent moves of the wh-element from Spec.C to Spec.C. By contrast, the subsequent cases of move <+wh> do not pied-pipe the CP *wie/die/wat/dat ze kuste* ('that she kissed'). The CP is stranded.

The explanation for this pied-piping asymmetry is quite simple by a re-interpretation of Ross' (1967) Left Branch Condition (Van Kampen 1997, 2004). The original NP complement (*jongen*) of the wh-element (*welke*) is pied-piped since as an NP it needs the D° *welke* as a case licenser, see (15).

- (15) a. *welke kuste zij [D t_{wh} [jongen]_{NP}]_{DP} ?
 which kissed she [D t_{wh} [boy]_{NP}]_{DP} ?
 b. [D welke [jongen]_{NP}]_{DP} kuste zij t_{<wh>} ?
 [D which [boy]_{NP}]_{DP} kissed she t_{<wh>} ?
 (which boy did she kissed?)

When case targets N°, i.e. in languages with morphological case (Slavic, Latin), the complement NP does not need the D° element, the Left Branch Condition does not hold and subextraction of the D°_{<+wh>} is possible. In languages without morphological case on the N° (Dutch, German) case targets D° (as proposed by Lebeaux 1988: 242f) and pied-piping of the NP complement follows obligatory. Note that it is not the richness of the morphological case-paradigm, mentioned in Ross (1967) and Uriagereka (1988), but the target position of the case-marking (either on D° or on N°) that quite naturally activates or deactivates the Left Branch Condition. In partitive constructions (*combien de livres/wat voor boeken* ‘what for books’) the preposition takes care of the case-licensing and pied-piping becomes an option.

Subsequently, there is an economy conflict between preserving major arguments (by pied-piping) or minimally moving only the D°_{<+wh>} by subextraction. When the Dutch child starts using complex wh-phrases, she first moves the D°_{<+wh>} alone, see (16)a. Only in a later acquisition step, after the age of 4, the entire wh-phrase is pied-piped, see (16)b.

- (16) a. welke wil jij [t_{wh} liedje] zingen? (Sarah 3;7)
 which want you [t_{wh} song] sing?
 (which song do you want to sing?)
 b. [welke verhaaltje] wil jij t_{wh} voorlezen? (Sarah 4;2.8)
 [which story] want you t_{wh} read
 (which story do you want to read?)

Adult Dutch allows subextraction in limited contexts (Corver 1990) that can be explained along the lines above (Van Kampen 2004). NP complements, though, are obligatorily pied-piped, as in (16)b. The order of acquisition step now shows that stranding is not learned, it comes for free. What is learned is that the remnant has to be pied-piped given certain licensing conditions, i.e. case-marking for NPs. These licensing conditions have to be learned. The enlarged options for subextraction in child language start in a period when licensing of NP by a D° is still optional. This perfectly fits the present proposal.

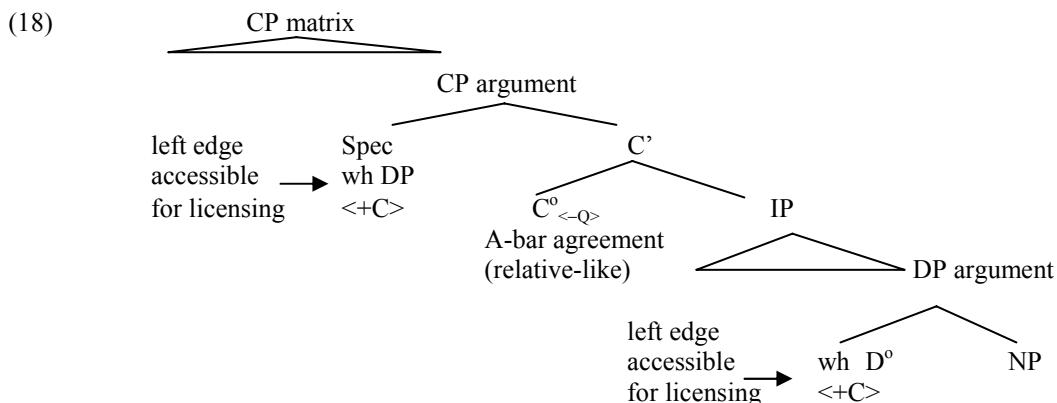
The same pied-piping story holds for the wh-phrase on the left branch (the left edge) of the CP. If the wh-phrase is on the left edge Spec of a CP marked as <+Q>, it cannot be extracted anymore. It is a <+Q> licenser for the CP<+Q>. Therefore in (17)a the entire CP is topicalized, pied-piping the whole clause *jij kent* (‘you knows’) along with the wh-phrase *wie* (‘who’). By contrast, the same wh-phrase on the left edge Spec of a CP<-Q> will not license that CP, will not activate the Left Branch Condition and must be subextracted, see (17)b.

- (17) a. [wie jij kent]_{CP<+Q>} weet ik niet t_{CP<+Q>}
 [who you knows]_{CP<+Q>} know I not t_{CP<+Q>}
 (I don't know who you are familiar with)
 b. wie_i denk je [t_{<wh>i} [dat ik ken t_{<wh>i}]]_{CP<-Q>} ?
 who_i think you [t_{<wh>i} [that I know t_{<wh>i}]]_{CP<-Q>} ?
 (who do you think I am familiar with?)

This shifts the problem. It may be that the wh-phrase in (17)b can be freely extracted, but why did it ever land in such an intermediate Spec.C? After all, the wh-movement into an intermediate Spec.C position cannot have been triggered by a target position C<+Q>, since the weak assertive *denken* selects a <-Q> complement. See for this “triggering problem” Lasnik & Saito (1984). There is a semantically relevant trigger <+Q>, but where is the local trigger given the C<-Q> in (17)b? My proposal runs as follows. There are two triggers, a syntactically simplified trigger <+C> and a more specified trigger

<+C, +Q>. The trigger <+C> requires that each A-bar pronoun <+D, +C> gets positioned in the first C-projection beyond the local predicate-argument structure it is in. This may seem a re-description of the contention that rules have to be local, but the position of the first C°-up is a clausal scope-position that has to be checked anyway as a C°<+Q> or a C°<-Q> in order to find out whether the wh-movement has to be local or (pro)long(ed). If the first C°-up is a C<+Q>, the movement triggering feature <+C> will be deactivated, say removed. Otherwise, when the first C°-up is a neutral head C°<-Q>, (see e.g. the C° of the weak assertive *denken* ‘think’ in (14)/(17)b), the movement triggering feature <+C> of the wh-phrase will not be deactivated and remains active.¹

Suppose there is this local movement to the first C-projection, see (18).



The CP is an argument of the matrix CP. The C° of the argument is <-Q>. The left edge of that CP argument is accessible and (after the first wh-movement) spotted as a <+C>. Hence, the <+C> movement reapplies to the CP argument. There is a left edge target/source overlap in Spec.C. Yet, this time the wh-pronoun will not pied-pipe its complement (the CP complement), since it does not license that CP<-Q> complement. My central point is that now all contributing factors in the reapplication of wh-movement have already been acquired by the child from more elementary constructions. These are (i) accessibility of information at the left edge, (ii) the movement up to the first A-bar position, (iii) the ± pied-piping factor, (iv) the A-bar agreement from the relative paradigm.

My point is of course not to rediscover Chomsky's (1973) Spec.C escape hatch or Ross' (1967) Left Branch Condition. I only argue that the ingredients (i) to (iv) for long wh-movement are already in place long before the child moves on to apply them in a combined fashion. So, my point is that long wh-movement is not learned, but follows as an implication from simple acquisition steps that are not a priori or innate, but rather previously discovered in more elementary and more frequent structures.

4. A-bar pronouns from Spec-head agreement

Chomsky (1973) argues that the long wh-movement had to be the outcome of a successive cyclic passage of the wh-phrase through the intermediate Spec.C positions. This successive cyclic passage is reflected in child Dutch as an appearance of A-bar pronouns in the intermediate C° head positions. It seems natural to derive these intermediate A-bar pronouns by means of a Spec.C-C° agreement as in Thornton (1990), Van Kampen (1997). The chain is formed by the Spec.Cs, whereas the spell-out of the C°s is a reflection of local agreement. For that reason, the C°s do not form a chain, cf. (14).

The following point is of some interest. Standard adult language evades the use of A-bar pronouns in the intermediate positions of long wh-movement. It restricts itself to the neutral C° head (complementizer) *dat* only. Child language, by contrast, applies the intermediate A-bar pronouns for a long time almost exclusively. See the numbers in (19) for Laura and Sarah.

¹ The resulting intermediate A-bar pronoun may easily violate the case requirement for upstairs matrix arguments. The present analysis predicts then that grammars expressing case in their A-bar pronouns (Russian, Polish, Hindi, German and Hungarian dialects) are unlikely to favor long wh-movements of the type we see here. They may rather use Partial Movement. De facto this seems correct (see Stepanov & Stateva 2006).

(19)		agreeing pronoun	complementizer <i>dat</i>
	Laura (between 7;2-9;0)	74	5
	Sarah (between 4;7-6;0)	12	0

Two things may be mentioned. First, long-distance questions appear quite late. Sarah's first long-distance questions are attested after the age of five. There are no long-distance questions attested in the speech of Laura before the age of seven. This is long after the paradigm of question *w*-pronouns and especially the paradigm of relative agreement has been firmly established in the speech of the child (Van Kampen 2007). Second, although the intermediate A-bar pronouns appear spontaneously, this does certainly not imply that they have to be learned. Let alone that their pattern is innate.

Below I will give all possible and attested variants of the agreeing pronoun. The rule of relative agreement seems to apply in all relevant cases.

Simple (non-complex) *w*-pronouns don't have an N-complement. The agreeing A-bar pronoun in intermediate position needs only to vary for < \pm animate> and < \pm oblique>, not for < \pm gender> properties of the antecedent. In (20) all examples that should be possible are given. They are all attested in child Dutch as indicated. The data are from the Van Kampen corpus (CHILDES) and diary notes 1993-1997.

- (20) a. *wie*<+animate> denk je *wie*<+animate> er in de auto rijdt? (Laura 8;3.8)
 who th. you who there in the car drives?
 (who do you think drives the car?)
- d. *wat*<-animate> denk je *wat*<-animate> ik ga zeggen? (Sarah 6;4.13)
 what th. you what I go say?
- c. *waar*<+oblique> denk je *waar*<+oblique> mijn handen zijn? (Sarah 4;10.20)
 where th. you where my hands are?
 (where do you think my hands are?)

The set of agreeing *w*-pronouns in (20) is not complete. Example (21) occurs as well.

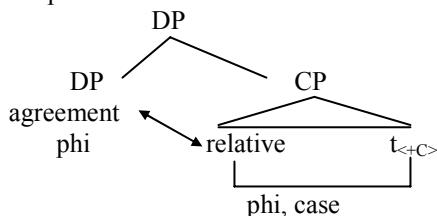
- (21) *wie* denk je *die* er jarig is? (Laura 9;1.4)
 who think you that there 'jarig' is?
 (who do you think has her/his birthday?)

In section 2 it was argued that spoken Dutch has the tendency to select *die* with antecedents that are semantically specified for <+animate>, cf. (11)b. It was shown to hold for relative agreement as well, cf. (11)a. I assume *die* in (21) to be specified for <+animate> as well.

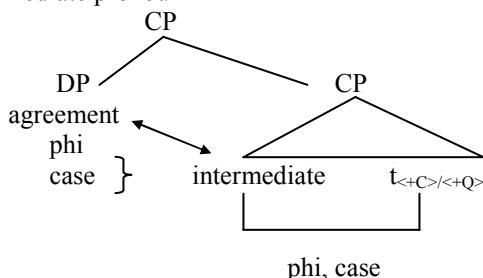
A parallel picture arises when full, complex, *wh*-phrases exhibit long successive movement. The agreement properties known from the relative paradigm, with all alternatives and irregularities, seem to reappear in the paradigm of the intermediate A-bar pronoun. These properties were already learned in a previous acquisition step.

The CP-adjoined DP closes further grammatical calculations in the CP with an agreement that seems to check the carry-over of information. Like the relative pronoun in (22)a (cf. (8)), the intermediate agreeing pronoun in (22)b agrees in phi-features with the moved *wh*-phrase.

- (22) a. relative pronoun



- b. intermediate pronoun



There is a difference between the top-labels in (22)a and in (22)b. My “CP-adjoined DP” covers up that difference, since it appears to be irrelevant for the A-bar agreement.

In (23) all possibilities of intermediate pronouns agreeing with a complex wh-phrase are given. Again, they are all attested in child Dutch as indicated. The variation of the intermediate pronoun shows that their choice is due to agreement, not to movement. Example (23)b with *dat* will further be left out of the discussion. It is a neutral C⁰ head generalized to all long distance questions in standard Dutch. Moreover it is not or hardly attested in the child data, cf. the figures in (19).

- (23) a. welke stift<+gender> denk je die<+gender> ik moet nemen? (Laura 7;8.18)
 which felt-tip think you that I must take?
 (which felt-tip do you think I must take?) (cf. (13)a)
- b. welk meisje denk je dat<-gender> ik leuk vind (child/adult Dutch)
 which girl think you that I like?
 (which girl do you think I like?) (cf. (13)b)
- c. in welk huis denk je waar<+oblique> jij woont? (Sarah 4;10.20/
 in which house think you where you live? Laura 7;7.10)
 (in which house do you think you live?) (cf. (14)c)
- d. op welke stoel denk je waar<+oblique> ik op zit? (Laura 7;10.1)
 on which chair think you where I on sit?
 (on which chair do you think I sit?) (cf. (14)d)
- e. op wie denk je op wie<+animate> Sarah verliefd is? (Laura 8;3.8)
 with who think you with who Sarah in love is?
 (with whom do you think Sarah is in love?) (cf. (14)e)
- f. welk meisje denk je die<+animate> ik een hand geef? (Laura 8;3.8)
 which girl think you that I a hand give?
 (which girl do you think I shake hands with?) (cf. (15))
- g. welk cadeau denk je wat<-gender> ik geef? (Laura 7;10.1)
 which present think you what I give?
 (which present do you think I will give?) (cf. (17))

The relative agreement paradigm excludes the cases in (24). These are unattested in child language.

- (24) a. *in welk huis denk je daar<+oblique> ik woon?
 (in which house think you I live?) (cf. (13)c)
- b. *wat voor boeken<+plural> denk je wat ik heb gelezen?
 what for books think you what I have read?
 (what kind of books do you think I have read?)

- c. *welke boeken<+plural> denk je wat ik heb gelezen?
 which books think you what I have read?
 (which books do you think I have read?)
- d. *welke villa<+gender> denk je wat ik ga kopen
 (which villa do you think I will buy?)

All intermediate obliques have to be <+wh> *waar* and cannot be <-wh> *daar*, as already predicted by relative agreement, cf. (9)c. The A-bar pronoun *wat* cannot agree with the plural *boeken* in relative agreement nor can it be spell-out of agreement in (24)b,c. The same holds for *wat* in (24)d. It cannot agree with the <+gender> noun *villa*. As a relative pronoun *wat* can only refer to a <-gender> noun. Therefore, *wat* can agree with the <-gender> noun *spelletje* in (25), cf. (23)g and (12).

- (25) a. wat voor spelletje<-gender> denk je wat ik wil doen? (Laura 7;9.27)
 what for game think you what I want do?
 (what kind of game do you think I want to do?)

The agreement solution also predicts correctly that the intermediate positions are never filled in by a repetition of the wh-phrase. The intermediates are for pronominal forms only. The examples in (26) are unattested.

- (26) a. *welke jongen denk je welke jongen ik leuk vind?
 which boy think you which boy I like?
- b. *welk huis denk je welk huis ik leuk vind?
 which house think you which house I like?
- c. *in welk huis denk je in welk huis ik woon?
 in which house think you in which house I live?

The agreement rule may extend to the preposition of oblique phrases, cf. (23)e. If the pronoun can express inherent case, as in (27)a,b (cf. (23)c), the preposition is not repeated. In (27)a *waar* reflects the locative. It corresponds with the antecedent *op welke school* ('at which school'). In (27)b *wie* is inherently marked for dative, which is possible in Dutch (ANS 1997: par. 5.5.8.2). However, if the intermediate *w*-pronoun cannot reflect oblique case, the preposition is repeated. In (27c) *wie* would not be an appropriate reflection for *op wie*, cf. (10)e (Van Kampen 1997 151f).

- (27) a. op welke school denk je waar Laura zit? (Laura 8;3.8)
 at which school think you where Laura sits?
 (at which school do you think Laura is?)
- b. aan wie denk je wie ik een brief schrijf? (Laura 7;9.2)
 to who think you who I a letter write?
 (to whom do you think I write a letter?)
- c. op wie denk je [op wie]_{co} Sarah verliefd is? (Laura 7;10.25)
 with who think you with who Sarah in love is?
 (with whom do you think Sarah is in love?)

The present approach suggests that the oblique preposition and its A-bar pronoun fit into the C^o head position as a complex head. See the brackets in (27)c above for an anomalous analysis. [P + A-bar features]_{co}. The nice outcome of the present approach is that relative agreement otherwise successfully explains the grammaticality of (27) versus the ungrammaticality of the three examples in (26).

This leaves me with a final difficulty. The intermediate A-bar pronoun *wie* is correct as a <+animate> spell-out of an intermediate A-bar pronoun for the phrase *welke jongen* and *welk meisje* in (28). Yet, relative agreement does not predict what spells out the also correct *die* for <+gender>/<+animate> antecedents, see the scheme in (13) (cf. (9)a/(11)a).

- (28) a. welke jongen denk je wie daar loopt? (Laura 8;3.8)
 which boy think you who there walks?
 (which boy do you think is walking over there?)

- b. welk meisje denk je *wie* ik een hand geef? (Laura 8;3.8)
 which girl think you who I a hand give?
 (which girl do you think I shake hands with?)

The main rule for relative pronouns is to select the *d*-pronoun if gender agreement is possible and to select the *w*-pronoun otherwise (Van Kampen 2007). Around the age of five when the long *wh*-movements and their intermediate A-bar pronouns begin to appear with some regularity in the speech of the child, the relative agreement pattern is already well established. The relative paradigm reappears for the intermediate pronoun agreement. Yet, the more specific gender agreement that determines the *d*-set weakens to an option under the more complex calculation of long *wh*-movement. Descriptively, the relative A-bar paradigm and the intermediate A-bar paradigm can now be stated as in (29).

- (29) A-bar agreement with a locally adjacent antecedent holds for relative pronouns and for the intermediate pronouns in long *wh*-movements.
- a. Relatives.
 Select a *d*-set pronoun if the antecedent has gender. Select a *w*-set pronoun otherwise.
 - b. Intermediates
 As relatives, or select a *w*-set pronoun if the antecedent is <+animate>, overriding the restriction on the *w*-set in (29)a.

The prediction in (29) excellently underlines that the order of acquisition steps is crucial for understanding the learnability of grammar.

From a somewhat broader point of view, one may notice that the present agreement proposal fits with Rizzi's (1996) *wh*-criterion. It also tallies well with the somewhat mysterious "doubly-filled Comp filter" (Chomsky & Lasnik 1977). Either the C^0 gets realized or the Spec.C, but not both. The answer may now run as follows. The trigger features <+C> and <+Q> on the C^0 head are erased when the specifier moves in its final landing side. The non-trigger <-Q> feature attracts phi-agreement. The *wh*-movement feature <+C> is not deactivated and consequently the *wh*-phrase <+D, +C> moves further.

5. Long-distance movement as an overlap of (cyclic) localities

In this paper I have defended the idea that all movement can be learned from example sets with the shortest steps possible, assuming that long-distance movement and island constraints follow from (i) left edge accessibility for the <+D> features, (ii) the minimality condition on movements <+C, +D> and (iii) (lack of) pied-piping.

It was argued that the child begins with subextracting the $D^0_{<wh>}$ of a complex *wh*-phrase. In a second step only, the NP complement is pied-piped. This order of acquisition step shows that stranding is not learned, it comes for free. Pied-piping is learned. Dutch allows subextraction in limited contexts (listed in Corver 1990). Hence, the child has to acquire the licensing conditions for the remnant, such as an obligatory D^0 context for case-assignment to NPs.

Subsequently, the child starts using long-distance *wh*-questions. In contrast to simple A-bar pronoun constructions, long *wh*-questions are fairly rare in the input. In the Van Kampen corpus (120 files of 45 minutes recordings, a total of 61.526 input sentences) only 4 long-distance *wh*-questions appeared in the speech of the mother. One cannot say that the child is "trained" on such structures. Nevertheless, they appear after the age of 5 and they take forms not present in the input. Four devices are to be combined by the child. (i) The short step trigger <+C>/<+Q> to $C_{<-Q>}$; (ii) The left edge accessibility (Chomsky's 1973 "escape hatch"); (iii) Obligatory pied-piping the NP complement; and (iv) The intermediate A-bar pronoun agreement.

A closer look reveals that all four devices are known from previous acquisition steps and have been acquired earlier from highly frequent simple contexts, see (31). The ages are a rough indication. Note that long-distance questions appear quite late in the speech of the child (after the age of 5). Long distance questions solicit the hearer's opinion and are by consequence dependent upon a theory of mind reading (Van Kampen 1997: 141). The young child is a formidable mind-reader, but making the systematic distinction between the inner and outer domain is a different matter and the corresponding grammatical devices come in later.

- (31) a. A-bar pronouns <+C> appear in initial position and they leave an argument gap: learned from all simple topic and question sentences. (before the age of 3)
 b. Left branch subextraction: learned from all wh-phrases. (before the age of 4)
 c. Obligatory pied-piping the NP. (before the age of 5)
 d. A-bar pronoun agreement: learned from relatives. (between 4-5)

Movement affects the left edge for reasons that were already known from the stranding constructions in (30)a above. So, movement of the wh-element and stranding the remaining constituent is old. The fact that the CP does not pied-pipe is old, since, in contrast to NP, the CP_{<-Q>} does not need a <+Q> licenser and it is only obligatory licensing that forces the learner into pied-piping.

The long wh-movements show how a short wh-movement to the nearest CP_{<-Q>} brought the A-bar pronoun into a new domain for a new short step. There happened to be an overlap of movement localities. This implies that lack of a locality overlap causes an island constraint for wh-movement, see Van Kampen (2008) for an example. That seems trivial, but the logical consequence is that syntactic islands need not be learned. They follow from any stupid non-overlap of movement localities. The learner discovers transparency as something already present in the system, the left edge extraction and the pied-piping triggers. The long movements follow without need to notice or learn island constraints.

References

- ANS (Algemene Nederlandse Spraakkunst) (1997) ed. by Guido Geerts et al. Groningen: Wolters-Noordhoff.
- Barbiers, Sjef, Olaf Koenenman & Marika Lekakou (2007) 'Syntactic doubling and the structure of chains', paper presented at the *TIN-dag* (Linguistics in the Netherlands), Utrecht, February 3.
- Chomsky, Noam (1973) 'Conditions on transformations', in: Stephen R. Anderson & Paul Kiparsky (eds.) *A Festschrift for Morris Halle* New York: Holt, Rinehart and Winston, 232-286.
- Chomsky, Noam & Howard Lasnik (1977) 'Filters and Control', *Linguistic Inquiry* 8.3, 425-504.
- Corver, Norbert (1990) 'The syntax of left-branch extractions', PhD. dissertation University of Tilburg.
- De Villiers, Jill, Peter de Villiers & Tom Roeper (2007) 'Wh-questions: Moving beyond the first phase', ms. Umass.
- Gutiérrez, Mariá Junkal (2006) 'Acquiring long-distance wh-questions in L1 Spanish, in: V. Torrens & L. Escobar (eds.) *The Acquisition of Syntax in Romance Languages*, 251-289. Amsterdam: John Benjamins.
- Kampen, Jacqueline van (1997) *First Steps in Wh-movement*. Delft: Eburon.
- Kampen, Jacqueline van (2004) 'An acquisitional view on optionality', *Lingua* 114-2004, 1133-1146.
- Kampen, Jacqueline van (2007) 'Relative agreement in Dutch', in: B. Los & M. van Koppen (eds.) *Linguistics in the Netherlands 2007*, 112-125. Amsterdam/Philadelphia: John Benjamins.
- Kampen, Jacqueline van (2008) 'The learnability of A-bar chains', in: Martin Everaert, Tom Lentz, Hannah de Mulder, Øystein Nilsen & Arjen Zondervan (eds.) *Theoretical Validity and Psychological Reality*. Amsterdam: John Benjamins.
- Lasnik, Howard & Mamoru Saito (1984) 'On the nature of proper government', *Linguistic Inquiry* 15.2, 235-289.
- Lebeaux, David (1988) 'Language acquisition and the form of grammar', PhD. dissertation University of Massachusetts.
- Postal, Paul (1966) 'On So-called Pronouns in English', *Monograph Series on Language and Linguistics* 19 ed. by Francis P. Dinneen, 177-206. Georgetown University Press.
- Rizzi, Luigi (1996) 'Residual verb second and the wh criterion', in A. Belletti & L. Rizzi (eds.) *Parameters and Functional Heads* Oxford: Oxford University Press, 63-90.
- Ross, John R. (1967) 'Constraints on variables in syntax', PhD dissertation MIT.
- Stepanov, Arthur & Penka Stateva (2006) 'Successive cyclicity as residual wh-scope marking', *Lingua* 116, 2107-2153.
- Strik, Nelleke (2006) 'L'acquisition des phrases interrogatives chez les enfants francophones', *Psychologie Française* 52, 27-39.
- Thornton, Rosalind (1990) 'Adventures in long distance moving: The acquisition of complex wh-movement', PhD dissertation University of Connecticut, Storrs.
- Uriagereka, Juan (1988) 'On Government', PhD dissertation University of Connecticut, Storrs.

Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)

edited by Jean Crawford,
Koichi Otaki, and Masahiko Takahashi

Cascadilla Proceedings Project Somerville, MA 2009

Copyright information

Proceedings of the 3rd Conference on Generative Approaches
to Language Acquisition North America (GALANA 2008)
© 2009 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-436-2 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Kampen, Jacqueline van. 2009. The 'Phased' Learnability of Long Wh-questions. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA 2008)*, ed. Jean Crawford, Koichi Otaki, and Masahiko Takahashi, 127-138. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2313.