THE ACQUISTION OF FIRST ORDER CP AND DP RECURSION*

Julianne Doner
University of Toronto

1. Introduction

In adult speech, we are able, in principle, to produce infintiely long sentences by embedding structures inside other structures of the same type. This type of structure is known as recursion. Recursion is thought to be a fundamental property of human language. As such, it is natural to assume it to be part of Universal Grammar (UG). The study of language acquisition can shed light on the nature of UG, since the only tools children start out with in language learning are those which are provided by UG. However, some studies have indicated that recursive structures are not acquired until relatively late (Pérez-Leroux et al. 2012, Roeper 2011, Roeper and Snyder 2005). Thus, studying the acquisition of recursion may allow us to determine whether it is acquired as if it is innate, or as if it must be learnt. Likewise, comparing the relative difficulty of the acquisition of different forms of recursion could reveal how exactly recursion is enabled by UG.

In this study, I found that recursive structures first appear as early as we expect in child speech. However, they also occur much less frequently in child speech than in adult speech. This indicates that, although the grammatical analysis of recursion is relatively easy, accounting for its early appearance, the production of recursive structures is difficult in terms of processing, accounting for the low frequency of use. This is consistent both with recursion being enabled by UG and the fact that children's use of recursive structures is somewhat delayed by some measures.

1.1 Recursion

Recursion is unique to humans, and is therefore argued to be part of the Faculty of Language, which separates human language from other (i.e., animal) communication systems (Hauser, Chomsky, and Fitch 2002, Roeper and Snyder 2005, Givón 2009, and Arsenijević and Hinzen 2012). This is part of the reason why it is expected to be part of UG and a component in the Language Acquisition Device (LAD). Again, if it is part of the LAD, then it should be an innate ability of children. However, Roeper and Snyder (2005) have noted that direct recursion, where a word category takes another element of the same

*I began this project in Ana Teresa Pérez-Leroux's graduate Language Acquisition and Linguistic Theory course, and it would not have gotten this far without her encouragement. Thank you! I would also like to thank the audience of the psycholinguistics reading group at the University of Toronto and the audience at the Canadian Linguistics Association, both of which gave valuable feedback.

Actes du congrès annuel de l'Association canadienne de linguistique 2013
Proceedings of the 2013 annual conference of the Canadian Linguistic Association

category as its complement, is restricted cross-linguistically in different ways. There must therefore still be some part of recursion which must be learnt.

Arsenijević and Hinzen (2012) argue that recursion can only occur across phases. Phases are units of deixis that must encode some sort of reference, such as to a proposition, event, or entity. Each phase will bear only one reference, since embedded phases will have intensional reference. When a phase is spelled-out, it leaves only the phase head and its specifier in the workspace. In this paper, I consider CP and DP recursion, since both C and D are phase heads and they both exhibit salient cases of recursion. There are many parallels between nominal and clausal structure; however, CPs are generally assumed to have more functional structure than DPs, and are therefore usually longer and more complex.

1.2 The Acquisition of Recursive Structures

There is evidence that children do not master the use of recursive DPs and CPs until relatively late in their development. For example, Phinney 1981 (in O'Grady 1997) ran an imitation task investigating the acquisition of embedding structures containing tensed and infinitival embedded clauses. He found that at the age of 3;7, children were still not imitating embedded clauses. Likewise, Diessel and Tomassello (2001) argue that children's first uses of what appear to be parataxis are instead modal operators, as in (1), since they are formulaic and lack productivity.

(1) You know, I see her.

Givón (2009) counters this by arguing that children do have the semantic and syntactic components of embedded clauses. These may, however, be distributed over several utterances.

There are also difficulties with the acquisition of complex DPs. Young children do not use recursive possessives, understand them, or are even able to imitate them (Roeper 2011). Pérez-Leroux et al. (2012) created a task designed to elicit recursive comitative DPs from children. In their study, only about a third of three-year-old children produced first-order embedding, and only about a quarter of four-year-old children. That these structures are acquired relatively late is surprising considering recursion is supposed to be a fundamental element of grammar and supplied to the learner through the LAD.

1.3 Mean Length of Utterance (MLU)

A recursive DP has a minimum length of three words, as shown in (2).

(2) a. DP [P DP]
b. Apples in bags.

Likewise, a recursive CP has a minimum length of four words, if we assume one-word subjects and that the embedded clause contains an intransitive verb, as shown in (3).

(3) a. DP V [DP V]
b. I said she jumped.

Thus, even if recursion is innate, we would not expect a child to produce recursive structures before they can produce three- or four-word utterances. The Mean Length of Utterance (MLU) of a child's speech is often used as an indicator of their development. We might expect children to begin producing recursive structures before their MLU reaches the minimum length of the construction, since the MLU is only the mean, and half of the child's utterances will therefore be longer than their MLU. On the other hand, most instances of a particular recursive structure will be longer than the minimum length indicated in (2) and (3). As such, the point when a child's MLU is roughly the same as the minimum length of the construction under consideration seems a reasonable expectation of when they might begin using that construction, if that construction is innate.

1.4 Order of Acquisition

When we consider the relative ordering of the acquisition of CP and DP recursion, there are three logical possibilities. First, CP and DP recursion may both be acquired simultaneously. This might occur if the possibility of recursion is not available until later in a child's development. Simultaneous CP and DP recursion would thus indicate that recursion is enabled by UG, but that its availability is delayed. Secondly, CP recursion might be acquired prior to DP recursion. This may occur if, for some reason, embedded clauses are the most salient form of recursion, and so the acquisition of embedded clauses somehow facilitates the acquisition of other forms of recursion. Finally, DP recursion might be acquired prior to CP recursion. This may result if recursive DPs are less complex than embedded clauses, and so appear earlier. Of these three hypotheses, the last would be the most neutral, especially because, as mentioned above, the minimum length of a recursive DP is shorter than the minimum length of a recursive CP.

2. Methodology

I used transcripts from the CHILDES database (MacWhinney 2000) of the child Naomi (Sachs 1983). There are transcripts of her available from the ages of 1;2.29 to 4;9.3.

The data was collected by manually scanning the transcripts for the constructions in question. In all cases, repetitions, portions of nursery rhymes or other memorized texts, imitations, and constructions which were incomplete or were adjacent to an unintelligible word were not included. As well, if the construction was distributed over more than one utterance, it was not included.

Recursive CPs were required to have at least two tensed lexical verbs in order to be included. Thus, nonfinite clauses, or clauses containing periphrastic tense constructions or auxiliaries were not included. The embedded clause must have been a complement to the matrix clause verb, headed by either *that* or a null complementizer. As such, relative and coordinated clauses were not included, nor were clauses with a question word as the complementizer. Finally, tag questions and *let's* constructions were also not included.

Recursive DPs consisted of a DP which contained a PP which, in turn, contained another DP. Thus, particle verb constructions and double object constructions (DOCs) were not included. Excluding DOCs was particularly difficult, as it was not possible to tell, in many cases, whether a bare DP was a recursive DP or a DOC with the beginning of the sentence elided. Whenever such a bare DP was preceded by a similar DOC, as occurs twice in (4), I counted it as a truncated DOC.

(4)	Mother:	You know what makes the milk brown?	[2;8.23]
	Naomi:	Uhuh.	
	Mother:	Yes you do.	
	Mother:	What comes out of the can in the refrigerator?	
	Naomi:	Milk.	
	Mother:	No. That we put in the milk.	
	Naomi:	We put in the…	
	Naomi:	**Glasses in the milk.**	
	Mother:	No. We don't put glasses in the milk.	
	Mother:	We put choc(olate).	
	Naomi:	**Chocolate in the m(ilk).**	
	Mother:	Chocolate. That's right.	

Otherwise, the bare DP was included. On the other hand, if a DP was preceded by an overt verb, and it was ambiguous between a DOC and a verb followed by a recursive DP, as in (5), it was not included. This includes any DP-PP sequences preceded by the verbs *have* or *got*.

(5) I tried one in my mouth. [3;2.10]

Finally, if the copula intervened between the DP and the PP, it was not included.

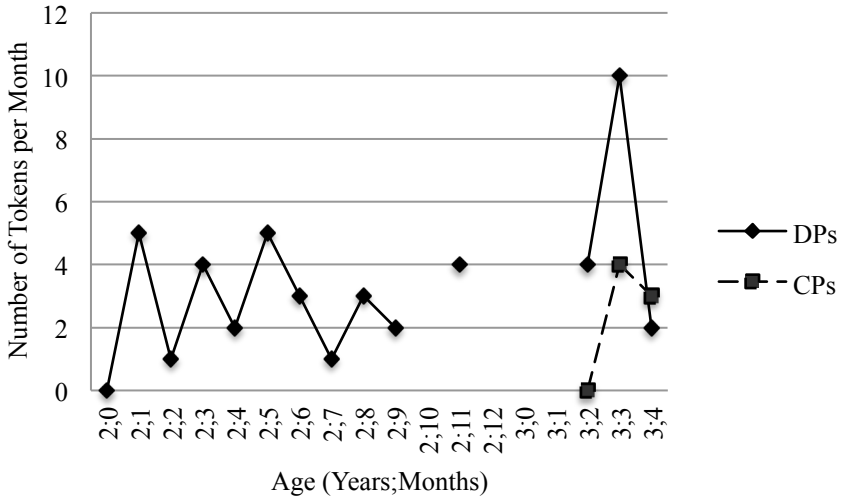
The adult data was coded the same way, with one exception. In child speech, we do not count imitations or repetitions of the adult speech, because we do not know whether it is an indication of the child's knowledge or the adult's. This does not apply in adult speech, as we know the adult has full competence. As such, when the adult repeated all or part of the child's utterance, it was still included in the tally.

MLU, the standard deviation of MLU, the mean length of the five longest utterances in a transcript (MLU5), and the total numbers of utterances were calculated by CLAN.

3. Results

The chart in (6) illustrates the development of Naomi's use of DP and CP recursion. The gaps in the data are from months in which there are no transcripts available.

(6) Emergence of CP and DP Recursion in Naomi's Speech[1]



Naomi begins using recursive DPs at the age of 2;1, and continues to use them consistently thereafter. She begins using recursive CPs at the age of 3;3.

3.1 First Use of Recursive CPs

Naomi starts embedding clauses with a tensed lexical verb as a complement to another tensed lexical verb at the age of 3;3.27. From the very start, Naomi's use of recursive complement clauses was productive. Naomi produced this construction with multiple different matrix verbs and subjects, as shown in (7).

(7) a. I don't think [you remember these little things]. [3;3.27]
b. This says [all the people live in here]. [3;3.27]

At the same age, she also produced multiple types of embedded clauses. In addition to the tranistive matrix clauses shown in (7), Naomi also produced the ditransitive matrix clause shown in (8).

(8) I whispered [that hurt] to you. [3;3.27]

[1] Note that there is not a consistent amount of data available for each month. There were no transcripts available for the months in which there is no data shown.

The embedded clause in (8) also cannot be concatenated, as it does not appear at the sentence edge. Additionally, the sentences in (7) and (8) contain a variety of subjects and verbs in their embedded clauses.

In Naomi's transcripts, adjunct clauses and embedded interrogative clauses are attested before embedded complement declarative clauses. I did not do a systematic search for them, but it is clear that she has productive use of these constructions by at least 2;11.13, over four months prior to her first use of embedded complement clauses. At this age, she produces embedded adjunct clauses with a variety of matrix verbs, both in the present and in the future, as shown in (9).

(9)	a.	I know where your shoes are.	[2;11.13]
	b.	I'll read two books if you'll feel better.	[2;11.13]
	c.	Even you sleep on this bed if you feel better.	[2;11.13]

The matrix clause can even be negated, as shown in (10).

(10)	I don't know what you drawed but you drawed +…	[2;11.13]

She also produces them with a variety of embedded clauses. In the examples in (9)-(10), we have three different embedded lexical verbs, in the past, present, and future. The form of embedding varies, as well. In (9)-(10), the embedded clauses are all interrogatives or *if*-clauses. In (11), we also have a temporal adjunct clause.

(11)	I'll read him a story before he goes to bed.	[2;11.13]

Thus, it's possible that Naomi acquired adjunct clauses and embedded interrogative clauses earlier than complement declarative clauses.

3.2 First Use of Recursive DPs

Naomi begins producing recursive DPs intermittently at the age of 2;1.1. Her first nine uses are shown in (12).

(12)	a.	Milk in it.	[2;1.1]
	b.	What in there?	[2;1.9]
	c.	Yup milk on it.	[2;1.25]
	d.	And sugar on it.	[2;1.25]
	e.	Berries on it.	[2;1.26]
	f.	Sugar on that?	[2;2.25]
	g.	All my cereal on it.	[2;3.0]
	h.	Sugar on it.	[2;3.17]
	i.	With the bee on it.	[2;3.17]

The context of (12i) makes it clear that it, at least, is a recursive DP, as shown below in (13).[2]

(13) Naomi: I want chocolate. [2;3.17]
Mother: okay.
Naomi: with the bee on it.

Naomi's use, at this time, is not very productive. All of the embedded DPs are pronouns, and all of her uses have the same semantic function: that of indicating location. Only two prepositions, *in* and *on*, are attested. Naomi continues this same pattern for several months. Productive use doesn't begin until 2;5.8, as shown in (14)-(15).

(14) Cook something with celery [?] in it. [2;5.8]
(15) A surprise for Raggedy_Ann. [2;5.8]

In both of these utterances, the embedded DP is phrasal, unlike in the utterances in (12), where it is always a pronoun. An additional preposition, *for*, is attested in (15), and two additional semantic functions are introduced: the embedded DP in (14) is comitative and the one in (15) is benefactive. The utterance in (14) is especially complex, since it exhibits second-order recursion. At this point, then, Naomi has unambiguously productive use of recursive DPs. She has used three different semantic uses of recursive DPs, and a variety of selecting and embedded DPs.

3.3 Frequency

In order to determine whether Naomi's use of recursion was adult-like, Naomi's frequency of use was compared to her mother's. I measured the mother's frequency of use of recursive DPs and recursive CPs in the file where Naomi was 3;3.27.[3] The results are shown in the tables in (16) and (17), below.

(16) The frequency of recursive CPs per 100 utterances

Age	Number of Utterances	Number of Recursive CPs	Frequency of Recursive CPs
3;3.27	331	4	1.2
3;4.18	182	3	1.6
Mother (3;3.27)	212	6	2.8

[2] The issue of ambiguity with small clauses was brought up by a member of the audience at the psycholinguistics group at the University of Toronto. This, unfortunately, was not something that was coded for. However, the data in (13) indicates that at least some of Naomi's early uses were indeed genuine recursive DPs. Any future work should also take care to code for potential small clauses.

[3] Note that the father was also present during at least part of this transcript, so that the mother didn't carry all of the adult conversation. That is why, in this transcript, the child has more utterances than the mother.

(17) The frequency of recursive DPs per 100 utterances

Age	Number of Utterances	Number of Recursive DPs	Frequency of Recursive DPs
2;1.1	190	1	0.5
2;5.8	414	5	1.2
2;8.14	402	3	0.7
2;11.13	309	3	1.0
3;3.27	331	6	1.8
3;4.18	182	2	1.1
Mother (3;3.27)	212	13	6.1

At her peak, Naomi is using recursive DPs in 1.8% of her utterances, whereas Naomi's mother uses them at the much higher rate of 6.1%. Likewise, Naomi uses recursive CPs in 1.6% of her utterances, at her highest rate, whereas her mother uses them at a rate of 2.8%. It therefore appears as though Naomi uses both forms of recursion less frequently than her mother. Naomi's rates are even somewhat high, since I only considered transcript files in which Naomi produced the type of recursion in question. On the other hand, Naomi uses both forms of recursion at a similar rate as the other once she has acquired them, indicating that they are equally difficult for her.

3.4 Mean Length of Utterance

Recall that a recursive DP has a minimum length of three words, as shown in (18).

(18) DP [P DP]

As shown in Table 3, below, when Naomi begins producing recursive DPs, her MLU is 2.137. When Naomi begins using recursive DPs productively, at 2;5.8, her MLU is 3.647, a higher amount than we predicted. However, at this point, she is using full DPs, which are normally longer than one word, and so the minimum length of the structure is longer, and so these results are still within the range of what we would expect.

(19) Naomi's MLU at the age of recursive DP acquisition

Age	MLU	σ	MLU5
2;1.1	2.137	1.1624	4.0
2;5.8	3.647	1.814	9.4

A recursive CP, on the other hand, has a minimum length of four words.

(20) DP V [DP V]

When Naomi first begins using recursive CPs, at the age of 3;3.27, her MLU is 4.323, only slightly higher than the minimum length of the utterance.

(21) Naomi's MLU at the age of recursive CP acquisition

Age	MLU	σ	MLU5
3;3.27	4.323	3.031	14

Thus, in terms of MLU, Naomi begins using recursive DPs and CPs about when we would expect.

3.5 Order of Acquisition

Naomi begins using recursive DPs productively over 10 months before she begins using recursive CPs. The binomial test (Snyder 2007) shows that there is a probability of $p < 0.000$ that these were acquired simultaneously.

4. Discussion

4.1 Competence and Performance

The timing of Naomi's first use of recursion (compared to MLU) indicates that the grammatical analysis of first-order recursion is relatively easy for her. However, there are several indications that she also finds first-order recursion difficult. First, PPs occur more often in contexts other than embedded under DPs. Second, Naomi's first uses are restricted semantically (they all indicate location) and structurally (they all consist only of a single pronoun). Finally, Naomi uses recursive structures less frequently than her mother. We can thus conclude that Naomi has grammatical competence in first-order recursion, but she doesn't have adult-like performance.

4.2 Phases

Recall that Arsenijević and Hinzen (2012) argued that recursion must occur across a phase boundary. As such, any recursive structure should also contain multiple phases. Multi-phase utterances should be easier to construct than single-phase utterances of the same length, since the child has only a portion of the utterance in the workspace at any given time. The difficult part about phases would be re-assembling them. If multiple phases are built at different times, but pronounced consecutively, they must be re-assembled at some point prior to pronunciation. I assume that this process is enabled by UG. Presumably, previously spelled-out phases of the same utterance must be stored in short-term memory while new ones are being built. This can then explain why pronouns occur earlier; they are phase heads, and thus do not get spelled out, but remain in the workspace and do not need to be stored. Thus, it seems likely that difficulties in producing recursive structures arise from the processing challenges associated with storing and re-assembling phases. Further evidence for this analysis comes from the fact that, when Naomi omits obligatory elements, producing

ungrammatical recursive structures, she omits them in the embedded contexts, as shown in (22)-(23).

(22) [I think [[your pocket] has [some money] in __]] [a little]. [3;3.27]
(23) [I like [the brown from [the __]]]. [3;3.27]

In the case of first-order recursion, an utterance consists of only two phases, so the difficulty of storing and retrieving phases may be trivial. However, it seems likely that this task becomes much more difficult in terms of processing as more levels of recursion and their resulting phases are added to an utterance. This may also explain why even some adults seem to avoid higher orders of recursion (Pérez-Leroux et al. 2012). If this is the case, then although recursion is supplied by UG, concatenation would still be easier because a portion of the structure does not need to be stored while the rest is built. What is difficult about recursion is storing and re-assembling the embedded phases.

4.3 Summary

Naomi begins using recursive structures about the time we would expect when compared to her MLU, and any difficulty she exhibits with recursive structures is consistent with the processing challenges of multi-phase utterances. Thus, the acquisition of recursive structures is consistent with it being enabled by UG. The differences in the acquisition of recursive CPs and DPs can be explained by their relative length and complexity, since recursive DPs, the simpler construction, are acquired earlier.

A question that remains is whether different forms of recursion are easier or harder than others. Most of the recursive DPs Naomi produced are formed with PP adjuncts, whereas, on the other hand, the only recursive CPs we considered were arguments. It is possible that recursion within adjuncts is easier than recursion within complements, as suggested by the appearance of CP adjuncts earlier than CP complements in Naomi's speech. Arsenijević and Hinzen (2012) note that the intensionality effects that result from recursion are weaker or even non-existent in adjunct clauses, which may be related to their relative ease or difficulty. As well, both the recursive CP and DP constructions we considered consisted of indirect, rather than direct, recursion. Comparing the relative difficulty of different forms of recursion could give us important insights into what, exactly, makes recursion difficult.

References

Arsenijević, Boban, and Wolfram Hinzen. 2012. On the absence of X-within-X recursion in human grammar. *Linguistic Inquiry* 43:423-440.

Diessel, Holgar, and Michael Tomasello. 2001. The acquisition of finite complement clauses of English: A corpus-based analysis. *Cognitive Linguistics* 12-2:97-141.

Hauser, Marc D., Noam Chomsky, and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298:1569-1579.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

O'Grady, William. 1997. *Syntactic Development.* Chicago: The University of Chicago Press.

Pérez-Leroux, Ana, Anny P. Castilla-Earls, Susana Bejar, and Diane Massam. 2012. Elmo's sister's ball: The problem of acquiring nominal recursion. *Language Acquisition* 19:301-311.

Roeper, Thomas. 2011. The acquisition of recursion: How formalism articulates the child's path. *Biolinguistics* 5:57-86.

Roeper, Thomas, and William Snyder. 2005. Language learnability and the forms of recursion. In *UG and External Systems: Language, Brain and Computation*, eds. Anna Maria Di Sciullo and Rodolfo Delmonte, 155-169. Amsterdam: John Benjamins.

Sachs, Jacqueline. 1983. Talking about the there and then: The emergence of displaced reference in parent–child discourse. In *Children's language, Vol. 4,* ed. Keith E. Nelson, Hillsdale, NJ: Lawrence Erlbaum Associates.

Snyder, William. 2007. *Child Language: The Parametric Approach*. Cary, NC: Oxford University Press.

Yang, Charles. 2010. Who's afraid of George Kingsley Zipf? Manuscript.