Missing Heads in Child Language

Teun Hoekstra, Leiden University & Nina Hyams, UCLA/OTS, Utrecht

This paper provides a unified account of root infinitives, determiner drop and null subjects in early stages of acquisition. The central hypothesis is that all three phenomena result from the optional underspecification of the functional head Number. The hypothesis allows us to capture on the one hand the crosscategorial generalization across nominal and verbal structures, and on the other, the crosslinguistic distribution of the root infinitive phenomenon. While the underspecification of Number determines the relevant drops in the syntax, the relationship between grammar and discourse is responsible for the difference between children and adult systems with respect to the exploitation of this option.

1. Introduction

In this paper we have 3 objectives: to provide a unified account of root infinitives, missing determiners and null subjects in early language. These properties are exemplified in (1) to (3). These phenomena occur in languages in which in the adult form finiteness, overt pronouns, and determiners are obligatory. Yet, in the child’s grammar these are only optionally expressed at the relevant stage of development. Our second goal is to account for the cross-linguistic variation observed with respect to children’s use of root infinitives, which we discuss below; and finally, to tease apart the effects of grammatical and pragmatic principles during what Wexler (1994) has called the "Optional Infinitive stage."

1. Root infinitives

- Papa schoenen wassen. (Dutch) b. Pas manger la poupée. (French)
  Daddy shoes wash
  not eat the doll
- Eve sit floor. (English) d. Thorstn Ball haben (German)
- Thorstn ball have

2. Null Subjects

- Want look a man. b. Veux pas lolo. (French)
  Want not water

3. Missing determiners

- Niekje ook boot maken. (Dutch)
  Niekje also boat make
- Open door. (English)

There are several reasons why we think that these phenomena are related:
- First, in each case the adult-like obligatory expression of some functional head -finiteness, subject pronoun, determiners- is suspended in the child’s grammar.
- Second, these properties of early language seem to co-occur in real time development (cf.

---

1 This is a modified version of a paper presented at WCCFL 14, which will appear in the proceedings of that conference. This research was partially supported by a UCLA Academic Senate grant (1994-1995) and a visiting scholar grant by the Netherlandic Society for Pure Scientific Research (NWO) to N. Hyams (Fall 1995).
- Third, these heads have in common that they are all points at which the sentence may be anchored into discourse. Finiteness places the event or state denoted by the verb at a time relative to the time of discourse; definite determiners pick out entities which are familiar in the sense of Heim (1983), that is, discourse referents; and subject pronouns may be deictic (cf. Hyams 1994). More specifically, finite I marks temporal specificity by placing the event described by the verb at a specific interval of time relative to the reference or speech time, while definite determiners and pronouns marks nominal specificity by referring to a specific (i.e. presupposed or familiar) entity. Thus, we will refer to this stage as the **Optional Specificity Stage**.

The theoretical and developmental proximity of these optional 'drops' points to them being one and the same phenomenon. Yet there is currently no unified account. Moreover, the fact that these heads have a parallel pragmatic function suggests that their apparent optionality in early grammar (in contrast to the adult grammar) is an effect of pragmatic principles. We will propose an account in terms of **underspecification**, which we define as in (4). Specifically, we will argue for the principle in (5).2

4. The value of a functional head is not grammatically determined.
5. The functional head Number is optionally specified in early grammar.

Since Number is represented both in the nominal and in the verbal system, its underspecification induces cross-categorial effects. Pro-drop and determiner drop result from an underspecification of Number in DP, while RIs arise when Number is underspecified in IP. As we will show below, the assumption that Number is at play in the RI-phenomenon will also accounts for the linguistic variation which is found with respect to this phenomenon in early grammars.

### 2.0 Underspecification of Number

Let us begin by summarizing our assumptions on clause structure which are relevant to the present discussion. We assume, following proposals in Johnson (1990) that Person, Number head their own projections, which are hierarchically arranged as in (6). (See Johnson for justification of this arrangement in Germanic).3

6. PerP NumP TnsP VP

---

2 There are other underspecification accounts of the data we are concerned with, for example Guilfoyle (1984) and Wexler (1994), who argue that T(ense) is an underspecified head in early grammar, and Clahsen, Eisenbeiss and Penke (to appear), who argue that AGR is underspecified. As we will show, these accounts fail on empirical grounds.

3 For the purposes of this discussion we can ignore further functional projections such as AgrOP.
We also adhere to a strong universalist basis, as per Chomsky (1992), according to which
languages do not vary with respect to absence or presence of functional categories, but rather
the variation comes about as a function of the morphological extensions for the values of the
different categories, which may yield different word order patterns. In particular, we shall
show below that inflectional paradigms can be differentiated in terms of which values of which
heads are directly represented in the morphosyntax. The range of variations we will consider
is plotted out in (7). In this chart, m stands for 'marked in the morphosyntax'.

<table>
<thead>
<tr>
<th></th>
<th>Person</th>
<th>Number</th>
<th>Tense</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>a.</td>
<td>m</td>
<td>--</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>b.</td>
<td>0</td>
<td>m</td>
<td>0</td>
<td>--</td>
</tr>
<tr>
<td>c.</td>
<td>0</td>
<td>0</td>
<td>m</td>
<td>--</td>
</tr>
</tbody>
</table>

Let us flesh out this table by looking at the inflectional paradigm of the finite verb in
Dutch present tense, given in (8):

8. Dutch *spreken* 'speak'
   - Ik spreek
   - Jij spreekt
   - Hij/zij/het spreekt
   - Wij spreken
   - Jullie spreken
   - Zij spreken

The morphosyntactic marking shows three distinct extensions: 0, -t, and -en. There is no
person distinction in the plural, so -en marks only Number. Making the assumption that first
person may just be unmarked, the -t extension found with second and first person singular
marks singular Number, rather than Person. There is no extension for Tense either, which
gets the unmarked value interpretation, i.e. non-past. The paradigm thus instantiates a b-type
language.

English, as in (9), is like Dutch, if we follow Kayne (1989) in assuming that *you* is
grammatically plural: first person singular is unmarked, -s marks singular Number, while
plural is expressed by the bare form.

9. English *sing*
   - I sing
   - you sing
   - he sings
   - we sing
   - you sing
   - they sing

In contrast, the Italian verb displays Person distinctions, as shown in the paradigm in

---

*If -t is indeed a marker for singular number, one might expect it to also mark singular past
tense verbs parallel to the marking of plural -en on past tense forms, cf. Dutch *wandelen* -en
'walk-past-plural'. Yet, the singular past tense forms have no overt extension for number, cf.
*wandel* -en 'walk-past'. This is parallel to the absence of number marking on English past tense
verbs, cf. *walked*, *walked*. Blocking of morphological extensions of higher functional
heads by overt lower heads is a general phenomenon in Germanic as observed by Johnson
Italian, then, instantiates (7a): there is again no marking of Tense, as in Dutch and English, which again yields present tense as its value. For the purposes of this discussion we can remain neutral on the question whether Italian verbs are marked for Number. If there is marking for Number, it exists only in the plural. Like Italian are Catalan and Spanish.

10. **Italian** *parlare* 'speak'

   io *parlo*  
   tu *parli*  
   lui *parl*  
   noi *parliamo*  
   voi *parlate*  
   loro *parlano*

French is somewhat different, which may be the reason for its non-pro-drop status. Although the paradigm marks person distinctions in the plural, it does not in the singular, as is shown in (11). French thus constitutes a mixed system.

11. **French** *parler* 'speak'

   [je *parl*]  
   [tu *parl*]  
   [il *parl*]  
   [nu *parl*]  
   [vu *parle*]  
   [il *parl*]

Let us now turn to the distribution of the RI-phenomenon across child language. Although RIs occur in many child languages, they are not universal. As reported in Sano & Hyams (1994), the RI phenomenon does not really show up in the Romance pro-drop languages (Italian, Spanish, Catalan); the average rate of RIs in these languages is 6%. Also, as reported in Sano (1995), Japanese does not show an RI-effect. He reports that there were no root non-finite forms in the data of five Japanese children (ages 1;11 to 2;11). The languages where we do see a robust RI-effect are the Germanic languages such Dutch and German, and French, with rates ranging from 26% to 78% depending on the particular child and the particular language (cf. Guasti 1994; Pierce 1992; Weissenborn 1990; Weverink 1989). It seems to us that the correct descriptive generalization is that given in (12).

12. RIs are produced only by children acquiring the b-type languages.

The b-type languages are those which in their adult form show only an obligatory Number specification. This generalization is captured under our hypothesis that RIs result from an underspecification of Number in the early grammar.

To see how this is so, consider first Dutch and English as examples of b-type languages. In these languages, when Number is left unspecified, there is simply no finite morphology, and hence the infinitive surfaces. The virtual absence of RIs in pro-drop languages follows as well, since in these languages the verb will always carry Person marking. And as we will discuss further, root non-finite forms do not show up in Japanese, which always expresses Tense.

The inclusion of two languages in the RI-group requires further comment here, viz. French, which has, as we said, a mixed system, and yet, it patterns with the b-type languages, and German, which would appear to be an a-type language in that the verb inflects for Person, but nevertheless German child language shows the RI-phenomenon. We shall discuss these languages in turn, but we first have to consider one additional point of general relevance,
which we might call *The avoid plural phenomenon (APP)*.

It seems to be generally true across languages that young children have productive control of agreement for singular persons, but fail to use plurals. For example, Pizzuto and Caselli (1992) report that between the ages of 1:4 and 1:9 the three children in their study productively used the singular forms; plural forms did not appear until much later. Grinstead (1994), based on examination of three Catalan and three Spanish speaking children, also isolates an early stage (1;7 to 2;2) during which the children have only singular verb forms (predominantly 1st and 3rd persons; second person is rather infrequent - a point to which we return). Grinstead's interpretation of this result is that in the absence of contrasting plural forms, Number must be assumed to be unspecified. Person, on the other hand, is marked from the earliest stages. The APP is also reported for French children (Ferdinand 1994) and German speaking children (Clahsen et al. 1994).

Given the APP, the question of how French should fit in the total picture is again straightforward: as plural forms do not take part in the child's system, we are left with the singular forms, which, as we showed in (11), show no person distinctions. Thus, early French qualifies as an instantiation of type-b and RIs are therefore expected. This expectation is correct. (cf. Pierce 1992; Kramer 1993)

Let us now turn to German. The adult agreement paradigm is given in (13):

13. German *spazieren* 'walk'
   
   | ich spazier(e) | wir spazieren |
   | du spazierst  | ihr spaziert  |
   | er/sie spaziert | sie spazieren |

German shows Person agreement, so our hypothesis wrongly predicts the absence of the RI-phenomenon in early German. More can be said, however. At the early stages German children do not have the plural forms (Clahsen et al. 1994), as noted above. Nor do they have the 2nd person singular affix *-st*. In fact, Clahsen et al. use the acquisition of the 2nd person affix as criterion for acquisition of the head AgrS, which they argue is initially absent. Thus, up to the point that *-st* is acquired, the child's system is like Dutch, as in (8); children use 0, (*e*) and *t*. Modifying slightly Clahsen et al's idea that Agr is underspecified in early German, we might say that although Person is a specified head in adult German, German children initially misanalyze their language as being Dutch-like, i.e. Personless but with a specification for Number. But since in the early grammar, Number is optionally specified, as per (5), when Number is left unspecified, RIs surface. We therefore predict that once the children acquire *-st*, that is, they learn that Person is a specified head in German, the number of RIs should drop to Italian-like levels. Clahsen et al. and Duffield (1992) confirm that with the acquisition of *-st* there is a dramatic decrease in the number of RIs. German children behave as expected under our hypothesis.

This may not be the end of the story, though, as one might raise the further question of why Italian don't children initially ignore the Person agreement, and hence optionally use RIs. Given that this is not the case, we are led to assume that not just any F-cat may be underspecified, but rather that if Person agreement is found in the target language, children should converge on that instantaneously. So, what is it about German that misleads the child or allows her to overlook the fact that the adult language has Person agreement? We offer two pieces of speculation here. First, adult German does not seem to be well-behaved for a
language with Person inflection. As Platzack and Holmberg's (1989) study shows, Person agreement is an adequate predictor of V-raising to a relatively high position (let's say, to PersP in terms of the configuration in (6)). Yet, this expected correlation does not hold for German. With respect to V-raising, German rather seems to correspond to a b-type language, like Dutch. Second, Person inflection would appear to correlate with the phenomenon of null-subjects of the Italian variety, a property which children apparently acquire quite early in Italian (cf. Rizzi 1994). Adult German, however, does not have null-subjects Italian-style: rather, null subjects seem to instantiate Topic-drop of some sort (cf. Rizzi 1994 for discussion and references), but no genuine null-subjects licensed by rich agreement. So, neither postverbal null subjects (i.e. following the finite verb in V-second position) nor embedded null subjects are allowed in German. Given the absence of such correlations with Person agreement in the adult system, the child may decide on that basis that she is in a b-type language, and hence ignore the Person inflection. It is interesting to note, as documented in Duffield (1992) and Hamann (1994), that the acquisition of Person agreement (i.e. the -st form for second person) appears to yield an intermediate stage in which argumental pro-drop is found, i.e. null subjects occur in postverbal position in the manner of a true pro-drop language. This stage is short-lived and basically disappears after sentential embedding becomes robust.

A further point to make concerns a difference in the agreement systems of the Romance pro-drop languages and that of German. In a system with Person agreement, one of the extensions may be a 'zero-form, i.e. one person may be the unmarked form. In the German system, first person singular is unmarked, while in the Romance languages, the third person is the unmarked form, as shown in (14):

14.  

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>m</td>
<td>German</td>
</tr>
<tr>
<td>m</td>
<td>u</td>
<td>Italian</td>
</tr>
</tbody>
</table>

This difference may be important for the following reason: in both systems, there is a formal opposition. However, an extension for 1st person must be Person, rather than Number, since if the extension marked singular number, it would also occur in the third person. An extension for third person (as in German) is compatible with both Number and Person marking (cf. -s in English, which does mark singular, not third person). Since second person reference is unreliable as a predictor, as such reference may be grammatically third person, cf. Spanish usted) or plural (cf. French vous, English you), the opposition between first and third plays a more vital role. The Italian system gives an unambiguous indication of Person agreement, while the German system does not. Whether this imbalance in the morphological oppositions is a crosslinguistically valid predictor of the Person Agreement correlations (i.e. pro-drop and V-raising) remains question for future investigation.

An important observation to make here is that in none of the languages just discussed does the verb bear an extension for the value of present tense: rather, it is through the specification of a higher functional head, say Person or Number, that Tense takes on the unmarked value of the tense opposition, i.e. present or non-past. This observation is of particular relevance when we consider the RI phenomenon. One hypothesis as to why children have RIs is that in the early grammar, in contrast to the adult grammar, T may be unspecified (or missing) (Guilfoyle & Noonan 1989; Radford 1990; Wexler 1994; Rizzi 1994). However, the underspecification of Tense hypothesis does not really distinguish the child’s grammar
from the adult’s system in these languages at the level of morpho-syntax; present tense is not morphologically expressed in the adult’s or child’s systems. So, not only does the underspecification of Tense assumption fail to provide a basis for the crosscategorial generalization that we seek to capture, since DP does not contain Tense, it is also an assumption which is hard to understand at the level of morphology—precisely, the realm where linguistic variation, including that between adult and child grammars, is supposed to be located.

Let us finally turn to c-type languages, such as Japanese. In these languages the verb inflects solely for Tense. For example, the finite form of the verb ‘see’ in Japanese is miru for all persons. Since these languages have no Number agreement, we predict that RIs should not occur. As we reported above, this is indeed the case for Japanese as Sano (1995) shows. The situation in Japanese is parallel to the Italian/Catalan/Spanish one; RIs do not surface in the pro-drop languages because Person must be expressed, and they do not surface in Japanese because Tense must be expressed.

The Japanese results are of particular interest because they bear on two alternative analyses of the RI-phenomenon, viz. the underspecification of Tense hypothesis mentioned above and the truncation analysis. If RIs resulted from an optional underspecification of Tense, we would certainly expect them to show up in early Japanese, as Japanese is the only language among those we have discussed which actually always marks Tense morphologically. But this is not the case. Neither can the Japanese situation easily be accommodated by a truncation analysis of the sort proposed in Rizzi (1994). According to Rizzi, RIs occur when the clause is truncated below the TP level. RIs will not occur in languages such as Italian in which the infinitive is forced to raise to a position higher than T, for example, to AgrS. The presence of the higher projection entails the presence of TP and hence excludes RIs. It is unclear, however, how such an account might be extended to a language like Japanese, in which neither finite nor non-finite verbs have extensions for AGR.

To conclude this section, we propose that in the early grammar Number and only Number may be optionally specified, as in (5). It is not the case that any functional category may be underspecified. Rather, if person (or tense) agreement is found in the target language, children should converge on that instantaneously. Only under the assumption that underspecification is limited to Number are we able to make the fine-grained crosslinguistic distinctions that seem to be required. (For discussion of why Number has this special status, see Hoekstra and Hyams 1995b)

---

5 One might expect that the Mainland Scandinavian languages would behave like Japanese on the assumption that the -er suffix marks present tense. Yet they show a robust RI effect. However, the overt marking of present tense would be surprising in a Germanic language. Rather, we would argue that the uniform -er suffixation in the Scandinavian languages is the result of the erosion of former Person and Number agreement markings. An indication that -er indeed does not mark present tense is the fact that it disappears in the presence of passive -s (cf. Norwegian jeg walger ‘I choose’, jeg walges ‘I am chosen’), which, however, combines freely with an overt past tense marker (cf. jeg valgte ‘I chose’, jeg valgtes ‘I was chosen’). There is therefore no incompatibility between Tense and -s which would account for the disappearance of -er. Thus we maintain that -er in fact instantiates Number even though it does not express a number distinction. On the disappearance of -er in the past tense, see note 4.
3. The Interpretation of Root Infinitives

Let us turn now to matters of interpretation. Before addressing the question of how RIs are interpreted, we lay out our assumptions concerning the interpretation of Tense in finite clauses. We assume that in a finite clause the event time in Tense is fixed relative to the speech time, which is itself represented in the C-domain, for example in the form of a deictic operator (Enç 1987; Guéron and Hoekstra 1994). The value of Tense fixes the event time as either prior to speech time, i.e. the past, or as identical to speech time, i.e. the present. Hence, there is an interdependence between C and T/V and it is through this interdependence that the event is situated. The temporal interpretation of a finite sentence is thus dependent on a (tense) chain (Guéron and Hoekstra 1994) which links T/V to C. When Tense is not morphologically expressed, as in the Germanic and Romance languages under discussion, it receives the unmarked [-past] value through the specification of the higher functional heads, Number and Person. This might be captured in terms of an extended chain, as in Chomsky (1986); that is, the specification of functional heads (eg. Number and Person) creates a chain between the C and the V/T (or between the spec of C and the Spec of V/T), as in (15).

15. OPi Ci NUMi [TP Ti [VP V]]

Thus, finiteness, in whichever head expressed, is instrumental in creating a link between C and T. If such a chain is established, T is a pronominal variable, bound by the operator in C.

What happens in RIs? We propose that the lack of specification of Number breaks the chain. In fact, no effects of a C-operator are visible in infinitives. For example, they do not occur with a that complementizer, nor do they ever undergo V-second. In RIs Tense is still pronominal, but no longer a pronominal variable, since no chain is established between C and T. This means that the interpretation of an RI is not fixed grammatically, but rather in the manner of a deictic pronoun, i.e. it is interpreted discursively. So, we take it that children make use of underspecification in the grammar, and this yields a direct discourse interpretation. This squares well with the general idea that children are more discourse-dependent than adults.

4. The nominal system

The analysis just outlined extends straightforwardly to the nominal system. We propose that D-drop (which includes the determiner drop and null subjects) in early language illustrated in (2) and (3), is the nominal analogue of RIs and that specificity is finiteness is the nominal domain.

To elaborate, we assume, following much current work, that the internal structure of nominal phrases is strongly parallel to that of clauses, i.e. CPs, with a determiner like the parallel to a complementizer like that. Between the D and the lexical noun, a number of functional categories must be assumed, including Number, as proposed by Ritter (1989) and Valois (1991). Thus, parallel to the clause structure in (15), we assume a DP structure as in (16).

16. OPj Dj NUMj [DP Xj [NP N]]
We leave the nature of X open at this point, noting that it is the nominal counterpart of T, which may be identified with the lower D in Szabolcsi's (1994) analysis of Hungarian.

Parallel to the C - T/V dependency within CP, expressed by finiteness, we have a dependency between D and X/N. The latter dependency is expressed by nominal 'finiteness', i.e. a specification of one or more of the intermediate functional heads (eg. Number). The D-X/N chain makes X a pronominal variable whose value is determined by an operator, i.e. specific (familiar or presupposed).

The parallelism between D and C is clear in such languages as Arabic, where N may move to the D-position in a way which is essentially analogous to V-to-C movement in clausal structures. Also, the interdependence of NUM and D is evident in the Romance languages where the determiner inflects for Number agreement, eg. French le fille, les filles. Our hypothesis predicts that the absence of a Number specification in the nominal system has the same effects on the D-N chain as it has in the clausal domain: the chain is broken. Morphosyntactically, this leads to the absence of a determiner, as in (3), as well as a lack of plural marking on N. The interpretation of the underspecified DP is similarly not grammatically determined via a D-X/N relationship. Rather, since X is not linked to D, it has the status of a pure pronominal, rather than a pronominal variable and as a pure pronominal, it is interpreted discursively.

These assumptions immediately carry over to pronouns. We assume the same basic structure for adult pronominal DPs as the one given in (16), with the difference that N may be empty. The effect of lack of Number specification then likewise leads to X being a pure pronominal. In effect, we have a radically empty DP, with the status of a free pronoun.

5. Why adults don’t drop

An immediate question that arises is why do adults not have the option of a direct discourse interpretation of T/V and X/N and hence use RIs and determinerless DPs. Limitations of space prevent us from discussing this issue in detail (see Hoekstra and Hyams 1995b). Briefly, we believe that this follows from a general pragmatic condition, formulated by Reinhart (1983) - her Rule I, which make discourse interpretations of pronouns unavailable if such interpretations are possible through binding. It is well-known that children are unaware of this bleeding relationship between grammar and discourse interpretation, allowing a wider use of discursively interpreted pronouns (the familiar principle B mistakes) (Wexler & Chien 1989). The option of direct discourse interpretation of functional heads, and hence RIs and D-drop (pro-drop and determiner drop), disappears with the development of the pragmatic system.

References


Hoekstra, T. and N. Hyams (1995b). The role of Number in verbal and nominal finiteness. [in progress].


