The role of input frequency in article acquisition in early child Swedish

Ute Bohnacker

This paper investigates patterns of article use in monolingual early child Swedish and in child-directed adult speech. Article omissions in the adult data are found to be more widespread than previously assumed, especially articleless, “bare” singular count nouns (e.g. *sten* ‘stone’ instead of *en sten* ‘a stone’) and article omissions in doubly determined nominals (e.g. *lilla tummen* (little thumb-the, ‘the little thumb’) instead of *den lilla tummen* (the little thumb-the; ‘the little thumb’). Such omissions in the input may arguably influence the course of acquisition. In the two children studied, an initial determinerless stage (1;3–1;7) is followed by a stage of optional articles (1;8–1;11). Targetlike article provision is reached at 2;0, which is early compared to most other Germanic languages. Definite enclitic articles (e.g. *-en* ‘the’ as in *sten-en* ‘the stone’) emerge at an earlier age and are produced at higher frequencies than indefinite prenominal articles (e.g. *en ‘a’* as in *en sten* ‘a stone’) and at an earlier age and at much higher frequencies than definite prenominal articles (e.g. *den ‘the’* as in *den lilla tummen* ‘the little thumb’). These child frequency patterns appear to replicate those of the adult caregivers. However, input frequency is argued to be an insufficient explanation for Swedish article acquisition, because of striking mismatches in child and adult article use in other areas, especially bare nouns. Investigations of child-directed adult speech are nevertheless important because they tell us what the immediate target looks like for the young child, which may be different from what linguists and reference grammars tend to assume.

1. Introduction and background

For a wide variety of languages, empirical studies have shown that articles tend to be omitted in very early child speech. When articles do emerge, acquisition is typically not instantaneous but gradual: children produce both determined and bare forms of the same noun (e.g. *a fish* and *fish*) in similar
syntactic contexts and with similar semantic and pragmatic functions. Acquisitionists have tried to account for the initial absence of articles and for their variable provision in a variety of ways.

Within the nativist generative paradigm – the paradigm that I have been working in – universal grammar (an innate faculty within general cognition but domain-specific to language) is assumed to generate mental representations of language structure. Minimally, language learners need to (a) set parameters, which will trigger discrete properties of grammar, and (b) learn lexical elements (free and bound morphemes) and the syntactic, semantic and phonological properties associated with them. Though some researchers have tried to attribute article omissions to parameter missettings, it is more widespread to explain article omissions via lexical learning, where the child may have a targetlike syntax, but is still learning the lexical elements and, importantly, the properties associated with them, a process that will be gradual and item-by-item. There may also be performance or processing deficits that need to be overcome in order to converge on the target (though often little is said about how exactly this is achieved). Alternatively, Roeper (1999, this volume) and Yang (2002) have proposed that children have co-existing and competing innate UG-constrained grammars. This would mean, for the nominal domain, that an overt-article grammar and a null-article grammar, or several more refined grammars, compete with each other, leading to variable production. In other syntactic domains, other sets of grammars will compete with each other. Depending on how well these grammars can parse the input, they get strengthened or weakened, and finally eliminated, until children converge on the adult distribution. Here, frequency may play a role when the child makes a choice between different mental representations (such as different parameter settings or weighting multiple grammars), but does not explain the construction of the actual mental representations, which are seen as UG-generated (cf. Roeper, this volume).

On any of the above generative approaches, input frequencies of linguistic forms can influence the course, speed and order of acquisition, and in the present paper I will investigate some aspects of this influence for the acquisition of articles in early child Swedish.

Articles and other determiners (demonstratives, possessives, quantifiers, etc.) and the grammatical features that linguists may attribute to them (definiteness, specificity, number, person, gender, etc.) are part of nominal phrases. A determined nominal phrase is generally thought to minimally have the syntactic structure in (1), where the determiner heads its own func-
tional projection (DP, determiner phrase) on top of the lexical noun phrase (NP). More elaborate structures have been proposed as well, but will not be discussed here.

\begin{equation}
\text{[DP \ldots [D D [NP \ldots [N N]]]]}
\end{equation}

There is, however, little consensus on whether bare, i.e. determinerless, nominals in adults (e.g. milk, fish) project to the DP level (with an empty or covert D), or whether they are simply NPs. Children’s bare nominals (which are often instances of article omission) could likewise be characterized as NPs only, and variability in children’s article provision as competition between a DP-grammar and an NP-grammar (see Kupisch 2004 for this line of reasoning).

During the 1980s and 1990s, children were often said to lack functional categories and projections (e.g. DP) in their mental representations until the age of ca. 2;0, at which this functional structure would mature in a genetically pre-determined manner, much in the same way as nonlinguistic processes like teething (e.g. Radford 1990a,b). However, other researchers were quick to point out that such theories of maturation made the false prediction that children should acquire articles roughly at the same time (2;0) across languages, while empirical studies showed just the opposite: individual children often differ in their rate of article acquisition, and cross-linguistically, articles have been found to emerge at very different ages and reach acquisition criterion at different ages for different languages. There are too many studies to be cited here, but to name a few, Lleó (1998, 2001) showed that both monolingual and bilingual infants start to produce articles earlier in Spanish than in German; Kupisch (2004, this volume) showed that both monolingual and bilingual children acquire articles earlier in French and Italian than they do in German, and in earlier work (Bohnacker 1994, 1997, 2003) I showed that articles are acquired earlier in Swedish than what has been reported for child English and German. Moreover, different types of determiners (e.g. definite articles, indefinite articles, demonstratives) are often found to emerge at different points in time within a language for the individual child, and the same goes for the point at which such morphological markers reach acquisition criterion. For these empirical reasons, theories of maturation of functional structure have largely been abandoned.

Alternative explanations try to relate cross-linguistic differences in article acquisition to individual and/or language-specific properties of the in-
put. Researchers of various theoretical persuasions increasingly realize the necessity of investigating the input that children actually receive. Recordings show that child-directed speech may diverge from adult-to-adult speech and from the ultimate target language, and sometimes there are surprising similarities (e.g., distributional patterns and frequency matches) between child-directed speech and early child production. Such findings are highly important, and as linguists we may have to revise previous assumptions about what the input and immediate target is for the child, and sometimes even revise our assumptions about the grammar of the language (variety) in question (cf. Bohnacker 1999 for Icelandic; Kupisch 2004 for Italian). I believe this renewed interest in input issues is a healthy and promising development.

If certain morphological markers, syntactic constructions and semantic/pragmatic functions are frequent in the input, will these markers, these constructions, these functions be the ones to emerge early in the child? Will they become productive early on? Will they also reach acquisition criterion early (whichever way that may be defined)? And will low-frequency items emerge late, become productive late, be acquired late? Existing studies tell us that such correlations are not always corroborated empirically—not all high-input frequency forms are acquired early for instance, which suggests that there must also be other factors at play.

Apart from determining the effects of statistical structure of the language on the order of acquisition, we may also ask what input frequencies actually tell us about the process of acquisition. In non-generative frameworks, closely corresponding frequency patterns between input and child production have been taken as evidence for stochastic learning, for an acquisition device that takes the statistical properties of the input to be the primary determinant of development, basically computing and matching the frequency of various elements in the input. What is frequent will become “entrenched” (cf. Langacker 1987), what is infrequent will not. The child is assumed to make lexically-specific associations, build a repertoire of learned strings (formulae) and use these to construct schemata and more general schemata (slot-and-frame), ultimately a type of pattern learning that is not domain-specific to language (usage-based, frequency-based, item-by-item acquisition models, cf. e.g. Lieven et al. 2003; Tomasello 2003; Theakston et al. 2004). Constructivists thus assume that the child constructs mental representations from the statistical structure of the input, in the absence of any innate language structure. Nativists, on the other hand, assume that mental representations of language structure are in place from the out-
The role of input frequency in article acquisition in early child Swedish (UG), but that their specifics are triggered by the input to which the child is exposed. (The input is also the source for lexical learning of morphemes and the properties associated with them.) On this approach, the target of acquisition is not knowledge of frequency; in fact, the statistical structure of the input is only an epiphenomenon and with little explanatory force (cf. Bley-Vroman 2002: 210, 212).

In generative frameworks, frequency correlations between input and child production have sometimes been interpreted as evidence for very early parameter setting (e.g. Wexler 1994; Hyams 1996: 94–97), or for early convergence on the adultlike weighting of multiple grammars. Mostly however, researchers have focused on how and why children go beyond what they hear in the input. The indisputable existence of frequency mismatches and children’s creative, novel and at times non-targetlike utterances point to a grammatical system that is underdetermined by the input available, serving as arguments against learning models that are purely frequency-based. Of course, lexical learning, learned strings that turn into schemata and frequency pattern learning are vital components of language acquisition, they can influence acquisition orders and thereby affect the development of linguistic knowledge, and I would venture to claim that most generative acquisitionists think so (e.g. Hyams 1992, 2005). However, I do not believe that it logically or empirically follows from this that linguistic knowledge can be reduced to knowledge of the statistical properties of language, and that children do not have access to a generative, i.e. a creative, combinatorial language component and to innate principles of language (cf. also Jackendoff 2005).

In this paper I will investigate the course of acquisition of articles in Swedish, by analyzing very early spontaneous production data from two monolingual Swedish-speaking children, Embla and Markus. In Sections 2 and 3, I will summarize relevant aspects of my earlier research (Bohnacker 1997, 2003) and for reasons of space must refer the reader to these works for details. Then I will take a slightly different look at Embla’s and Markus’s data and home in on the distribution patterns and frequencies of different kinds of articles and articleless nominals in child production vis-à-vis caregiver speech. I will focus on four potential input frequency effects and discuss whether these can help explain the course of acquisition: definite vs. indefinite article frequency, definite enclitic vs. definite prenominal article frequency, bare noun frequency and bare noun subtype frequency (Sections 4–7). I will argue that even though there are matches between input and early child production, we also find major discrepancies in fre-
quency, early evidence of rule productivity, and thus evidence for article acquisition being more than frequency-based statistical learning.

2. Articles in adult Swedish

Swedish has overt morphological definite and indefinite markers. The indefinite article used with singular count nouns is a free prenominal monosyllable, *en* ‘a’ or *ett* ‘a’ (e.g. (2a) *ett får* ‘a sheep’). There is no indefinite article for mass nouns (substances and collectives), abstract nouns, indefinite plural count nouns, and these therefore occur bare (e.g. *mjölk* ‘milk’, *enighet* ‘unity’, *ko-r* ‘cow-s’). The definite article is postnominal and enclitic and takes the form of a syllabic or non-syllabic suffix (e.g. (2b) *får-et* ‘sheep-the’, *ko-n* ‘cow-the’). Swedish also allows multiple morphological definiteness marking, so-called “double determination”, in nominal phrases that contain a demonstrative or adjective: here, a prenominal definite article and the definite enclitic article co-occur (cf. (2c–e)). Elements in the nominal phrase agree in number (singular, plural) and grammatical gender (common, neuter).

(2)  

a. *ett får*  
   sheep.SG.NEU  
   ‘a sheep/one sheep’  
   indefinite article

b. *får-et*  
   sheep.SG.NEU-the.SG.NEU  
   ‘the sheep’  
   definite article

c. *det här får-et*  
   the.SG.NEU here sheep-the.SG.NEU  
   ‘this sheep’  
   double determination

d. *dét får-et*  
   the.SG.NEU stressed sheep-the.SG.NEU  
   ‘THIS sheep’  
   double determination

e. *det svarta får-et*  
   the.SG.NEU black sheep-the.SG.NEU  
   ‘the black sheep’  
   double determination, adjective

e’. *svarta får-et*  
   black sheep-the.SG.NEU  
   ‘the black sheep’  
   single determination, adjective
It is less well known that Swedish definite nominal phrases with an adjective need not be doubly determined, but can carry only one definite marker: for certain classes of adjectives, there never is any prenominal definite article, and for other adjective classes, the prenominal definite article of the double determination construction is optional in many varieties, cf. *(det)* *svarta får-et* *(2e’)*. This fact is often ignored in theoretical accounts of Swedish and also, sadly enough, in studies of Swedish child language, which erroneously classify singly definite phrases such as *svarta får-et* across the board as non-targetlike prenominal article omissions.

Another often-ignored fact about Swedish articles is the widespread occurrence of “ungrammatically bare” singular count nouns without the prenominal indefinite article. Such bare singulars have a non-individuated reading, the type-properties of the noun take precedence over the individual (or token) properties (Bohnacker 2003: 201–202, 243–247), as illustrated in (3a).

This use is highly productive, and not – as some textbooks and reference grammars of Swedish would have you believe – restricted to a handful of idioms or fixed constructions (verb + bare singular, preposition + bare singular). (3) gives a selection of the bare singulars I noticed when listening to the radio and leafing through a newspaper magazine for only a few minutes.

\[(3) \quad \text{a.} \quad \text{return vi } _{\text{får}} \text{ och } _{\text{varg.}} \]
\[\text{now take we sheep and wolf}\]
\[\text{‘Now let’s talk about/take \{the/a/Ø\} sheep and \{the/a/Ø\} wolf/wolves.’}\]
\[\text{cf. } \text{return vi } _{\text{ett}} \text{ får och } _{\text{en}} \text{ varg.} \]
\[\text{now take we a sheep and a wolf}\]
\[\text{‘Now let’s take a/one sheep and a/one wolf.’}\]

\[\text{b. } \text{finns det } _{\text{varg}} \text{ däruppe?} \]
\[\text{exists it wolf there-up}\]
\[\text{‘Are there wolves up there?/Can the wolf be found up there?’}\]

\[\text{c. } \text{vill du } _{\text{påse?}} \]
\[\text{will you have bag}\]
\[\text{‘Would you like a bag?’ (cashier to customer)}\]

\[\text{d. } \text{jag har } _{\text{köpt}} \text{ ny tekanna. vill du } _{\text{se} \text{ den?}} \]
\[\text{I have bought new teapot want you see it}\]
\[\text{‘I’ve bought a new teapot. Do you want to see it (i.e. the teapot)?’}\]

\[\text{e. } \text{iförd } _{\text{jacka med fuskpåls och kapuschong}} \]
\[\text{wearing coat with fake-fur and hood}\]
\[\text{‘wearing a coat with fake fur and a hood’}\]
f. det är inte så lätt att vara _ svensk mamma.
   It is not so easy to be a Swedish mum
   själv har jag _ kurdisk mamma.
   myself have I a Kurdish mum
   ‘It’s not that easy to be a Swedish mother. I have a Kurdish mother myself.’

g. det är _ solig eftermiddag utanför Kungsholmens kyrka,
   it is sunny afternoon outside Kungsholmen church,
   det dof tar av vår.
   it smells of spring
   ‘It’s a sunny afternoon outside Kungsholmen church and spring is in the air.’

h. _ cykel går bra, men inte _ bil.
   bike goes well but not a car
   då sen så finns det ju _ taxi på ön.
   and then so exists it well a taxi on island-the
   ‘A bike works well/is fine, but not a car. And then there’s a taxi on the island.’

i. _ buss avgår klockan 9:00 från stationen.
   bus leaves clock 9:00 from station-the
   _ guide möter vid fågeltornet kl 10.
   guide meets at the bird-watchtower the clock 10
   ‘A bus leaves at 9 a.m. (i.e. There is a bus at 9 a.m.) from the station. There will be a guide meeting you at the bird-watchtower at 10 a.m.’

As (3d–g) show, bare singular count nouns can easily take modifiers such as postnominal prepositional phrases or prenominal adjectives (e.g. ny tekanna ‘new teapot’); “bare” refers to the fact that the noun occurs without an article. Bare singulars occur as predicates (e.g. vara _ svensk mamma ‘be (a) Swedish mum’; jag är _ lingvist ‘I am (a) linguist’) and as arguments (e.g. ha _ påse ‘have (a) bag’). Bare singulars often occur with copula verbs, as the object of thematic verbs and as the object of prepositions, but they also occur as the subject of existential, presentational and locational clauses, see (3b) and (3f–i). Swedish native speakers (and linguists) vary considerably in their acceptance of bare nouns, and when presented with bare singular count nouns they just produced themselves, some reject these in judgment tasks, indicating a mismatch in actual and perceived usage.
Many bare singulars occur in contexts where a determined nominal with an indefinite article could be used as well, e.g. (3c) vill du ha _ påse? (‘want you have Ø bag’) is just as acceptable as vill du ha en påse? (‘want you have a bag’), or (3d), where jag har köpt _ ny tekanna (‘I have bought Ø new teapot’) is equally acceptable as jag har köpt en ny tekanna (‘I have bought a new teapot’). The semantic difference between a determined and a bare noun is often very subtle; the bare noun emphasizes type properties, but does not rule out token reference, as shown in (3d), where a bare singular (ny tekanna ‘new teapot’) is the antecedent of a token-anaphor (den ‘it’, i.e. a specific teapot). Theorists are in disagreement as to whether bare singulars (and bare nouns in general) should be analyzed as DPs with an empty D, or simply as NPs, in which case Swedish could be said to have both a DP- and an NP-grammar. Even though there are no large-scale corpus studies of bare nouns as yet, Swedish appears to allow substantially more bare singular count nouns than closely-related languages such as German and English, and this may influence acquisition.

3. Recent insights about articles in child Swedish

Existing acquisition studies of Swedish articles have investigated tape-recorded naturalistic production data from the Stockholm-based Child Language Syntax Project of the early 1970s (Söderbergh 1973) and Strömqvist & Richthoff’s Gothenburg-based corpora of the 1980s/1990s (e.g. Strömqvist, Richthoff & Andersson 1993, partially available on CHILDES). Besides hard-to-come-by internal research reports, there have been few publications on Swedish article acquisition until recently. In Bohnacker (1997, 2003), I hand-searched the corpora of two of the Stockholm and Gothenburg children, Embla (1;8–2;1, 10 recordings) and Markus (1;3,19–2;0,25, 15 recordings), with additional data from Tor (2;3–2;5, 5 recordings), and investigated their nominal phrase development.

Interestingly, I found an initial no-article (no-determiner) stage, which, however, is short and can only be documented if data collection starts early enough. It is attested for Markus (1;3,19–1;7,25), but not for the other children for whom recording started at age 1;8 (Embla) or even later (Tor). It should be noted that the children’s speech in general is affected by processes of radical phonetic and phonological simplification, e.g. simplified syllable structure (CV, VC, CVC), consonant cluster reduction, vowel harmony, consonant harmony, stopping, final consonant deletion resulting in open CV syllables. These processes also affect nominals and may render
them as rudimentary forms that look like non-targetlike, determinerless bare nouns.

The no-article stage is followed by an optional-article stage (also attested for other children, cf. e.g. Lange & Larsson 1975; Svartholm 1978; Santelmann 1998). For Embla and Markus there is a rapid increase in article provision, the development is relatively fast and is reminiscent of how articles emerge at an early age in many children acquiring Romance languages (e.g. Spanish, Italian, French), but different from Germanic languages studied to date, such as English or German, where articles typically emerge late and reach adult levels late (cf. e.g. for German, Clahsen, Eisenbeiß & Vainikka 1994: 99; Eisenbeiß 2000; Kupisch 2004; Lleó & Demuth 1999; for Italian, Bottari et al. 1993/94; Crisma & Tomasutti 2000; Kupisch 2004, this volume; for Spanish, Lleó 1997; Lleó & Demuth 1999). Recent findings from Tromsø Norwegian (Anderssen 2006) suggest, however, that child Norwegian patterns with child Swedish.

Both Markus and Embla produce nouns with overt articles already at age 1;8 (corresponding to a MLU of 1.01 for Markus and 1.64 for Embla). Soon after, they produce sizable percentages of articles and by 2;0, both have reached adult provision levels (>90%). Markus’ article provision in obligatory contexts is shown in Figure 1 sample by sample in percent. The rightmost column shows a comparison with his adult caregivers.

![Figure 1](image)

*Figure 1.* Markus’ provision of articles in obligatory contexts, percent, vs. adult caregivers (based on Bohnacker 2003).
The role of input frequency in article acquisition in early child Swedish

Figure 2 displays the same information for Embla and her caregivers.

![Figure 2](image)

Figure 2. Embla’s provision of articles in obligatory contexts, percent, vs. adult caregivers (based on Bohnacker 1997).

Figure 3 illustrates Markus’ articles with common nouns, raw figures aggregated by stage (for sample-by-sample figures and a detailed analysis, see Bohnacker 2003). During his initial no-article stage (age 1;3,19–1;7,25) all nouns are bare, i.e. articleless (white columns). Bare nouns predominate also for the next few months (1;8,5–1;9,7), even though enclitic definite articles begin to appear at this point (checkered columns). At age 1;10–1;11,12, definite enclitic articles become more frequent, and prenominal indefinite articles (black columns) and prenominal definite article (grey columns) are emerging as well. Though not shown in the above chart (Bohnacker 2003), Markus produces high rates of definite articles before producing high rates of indefinite articles (67% (14/21) definite articles vs. 12% (10/85) indefinite ones at age 1;10,14; and 88% (60/68) definites vs. 62% (16/26) indefinites at age 1;11,12). Soon after, at 1;11,25–2;0,25, adult provision levels are reached for both types of articles and the distribution of articles in Markus’ productions closely matches that of the adults in the recordings, as shown by the rightmost column conglomerate in Figure 3.
Some examples of Markus’ and Embla’s use of articles are given in (4)–(6).

(4) Playing with cars and little toy people.
   *ADU:  jå  allting  finns kvar #.  
   yes everything is still-there
   *CHI:  anda gubbe-n.  
   other man-the
   ‘the other man’
   *ADU:  den andra gubben  ja  de bara en  gubbe i  bilen #.  
   the other man-the well it is only one man in car-the
   *ADU:  de ska  vara en gubbe till  ja.  
   it shall be a man more yes
   ‘There should be one more man.’
   *CHI:  gubb  ekaj #.  
   man.STEM [??]  
   (stem + unclear; perhaps imitation of ADU (en) gubb-e till)
   *CHI:  hu.  
   *CHI:  hm.  
   *ADU:  sa  du  att  du  e  blöt där  ja  du  ha-r  nog  kissat+...  
   said you that you are wet there yes you have well peed
*CHI:音响。
*ADU: há #。
*CHI: en gubbe. [...] a man
*ADU: jå.
*CHI: en gubbe. a man
*ADU: de fattas en gubbe #. it lacks a man ‘There’s a man missing.’

Markus (1;10,14)

(5)  a. *CHI: gubbe-n är glad. man-the is happy ‘The man is happy.’
    (Embla 1;8,2)
   b. *ADU: vad är det då? what is this then ‘What’s this then?’
   *CHI: en gubbe. a man ‘a man’
    (Embla 1;9,2)

(6)  a. *CHI: en bil. a car
    (Embla 1;8,2)
   b. *CHI: bil-en stå. car-the stand
    (Embla 1;9,0)
   c. *CHI: äh dén bil-en. no the car-the ‘no, THIS car’
    (Embla 1;9,0)
    ‘the cars, all the cars’
    (Embla 1;9,0)

In Bohnacker (1997: 68–69, 2003: 217–219), I argued that most children’s nouns with articles are productive uses of determiners (N+Art, Art+N, etc.) and not just frozen, unanalyzed wholes. I still take this view because Embla and Markus often produce the same noun both as a bare form and as a determined form with different articles, e.g. with a prenominal article in some
instances, with an enclitic article or a plural marker and an enclitic article in
others, as exemplified in (4)–(6). This suggests that the children have sepa-
rate entries for the noun and the determiners in their lexicons, and that they
are appropriately assigning a complex structure to the nominals with overt
determiners. What the abstract categories involved in this complex struc-
ture exactly are is up for linguistic discussion, and acquisitionists may dis-
agree about whether these abstract categories are given (innate) or con-
structed by the child; the point I want to make is simply that the children in
their productions in the corpora at hand appear to appropriately assign a
complex structure to the nominals that feature overt determiners.

Moreover, the children’s articles do not point to any semantic or prag-
matic deficiencies. Most of Embla’s and Markus’ early nominals with defi-
nite articles refer deictically or anaphorically to entities in the discourse,
and many of their nominals with indefinite articles are used in the function
of naming, as is often the case in early child language. In choosing between
definite and indefinite articles in these functions, Embla and Markus appear
to make hardly any errors, as far as one can tell from the transcripts.9 How-
ever, this need not mean that the children have learned all the semantic
functions articles have (cf. Kupisch 2004). Spontaneous production data
should of course be complemented by experimental data designed to test
semantic and pragmatic competence, e.g. Emslie & Stevenson (1981).

Interestingly, different morphophonological and semantic types of arti-
cles have been found to emerge at different times for Markus (Bohnacker
2003: 224–243), as had been hinted at earlier for other Swedish children
(e.g. Lange & Larsson 1975: 119, 1977: 6, 36; Svartholm 1978). As shown
in Figure 3, enclitics emerge in Markus’ speech (1;8) before prenominals
do (1;10), definite articles emerge (1;8) before indefinite articles do (1;10).
But they reach adult provision levels at the same time by 2;0. This stag-
gered development is interesting from an input frequency perspective (see
next section), and from a cross-linguistic perspective: is there something
about the semantics and/or function of a particular article that makes it easi-
ter to acquire? Or is there something about the morphophonological form of
a particular article that makes it easier to acquire (e.g. enclitic articles, met-
rical structure)?

I also found that different types of articles were produced with very dif-
ferent frequencies (Figures 4 and 5). For both Embla and Markus, most of
their articles are definite enclitics in singly definite nominals (Embla 78%,
Markus 51%). Indefinite prenominal articles are less common (Embla 10%,
Markus 35%). Double determination (definite prenominal + enclitic) is least common (Embla 10%, Markus 14%).

As the pie charts in Figures 4 and 5 show, the distribution of the children’s article types is very similar to that in the recorded input. It thus appears as if the children are indeed matching the frequency patterns of the input in their output. In the following sections, I will investigate this match in greater detail.
4. Potential input frequency effect 1: Definite vs. indefinite article frequency

Are definite articles the first articles to be produced in Swedish (Figure 3) because they are the most frequent article forms in the input? This has not been explicitly suggested but is implied in Bohnacker (1997). As the pie charts in Figure 4 and Table 1 show, Embla’s caregivers indeed produce many more definite articles (84%) than indefinite articles (16%).

Table 1. Definite vs. indefinite articles: Embla

<table>
<thead>
<tr>
<th></th>
<th>Embla’s caregivers (first 5 samples)</th>
<th>Embla (1;8,2–2;1,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite articles</td>
<td>84% (67/80)⁷</td>
<td>90% (161/178)⁷</td>
</tr>
<tr>
<td>Indefinite articles</td>
<td>16% (13/80)</td>
<td>10% (17/178)</td>
</tr>
</tbody>
</table>

Embla produces both types of articles already in her first sample, though raw figures are low and indefinite articles are rare (Bohnacker 1997: 66–67). We must therefore turn to Markus, whose corpus is much larger than Embla’s, for determining a developmental order. I hand-searched Markus’ samples and found that his caregivers produce definite articles slightly more frequently (55%) than indefinite ones (45%). This distribution appears to hold across the observation period. Caregiver and child frequencies roughly match, as can be seen from the pie charts in Figure 5 and Table 2.

Table 2. Definite vs. indefinite articles: Markus

<table>
<thead>
<tr>
<th></th>
<th>Markus’ caregivers (all samples)</th>
<th>Markus (1;3,19–2;0,2, all samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite articles</td>
<td>55% (615/1111)⁷</td>
<td>63% (235/373)⁷</td>
</tr>
<tr>
<td>Indefinite articles</td>
<td>45% (496/1111)</td>
<td>37% (138/373)</td>
</tr>
</tbody>
</table>

Markus starts to produce definite articles before indefinite ones (Figure 3), but reaches adult provision levels for both at the same time. This suggests that even though the frequencies of definite articles and indefinite articles as compared to each other match in child and parental speech, this by itself does not explain the acquisitional path (for a discussion of other, phonological, situational/interactional etc. explanations, see Bohnacker 2003: 232–247).
There is another issue that disturbs the otherwise impressively similar distribution of articles in child and caregiver speech (Figures 4 and 5), namely the issue of article omissions.

So far, I have concentrated on nominal phrases with overt articles. However, when we compare all the types of nominals that involve common nouns for Markus and his caregivers (i.e. all nominals, except for pronouns, proper nouns and name-like kinship terms), we find a strikingly dissimilar distribution. Markus produces predominantly bare nouns, with an average of 58% (567/927), whereas the adults produce 88% overtly determined nouns and only 22% (327/1487) bare nouns, as shown in Figure 6.

From the breakdown by age (Figure 3), we know that during the first months of recordings, Markus’ common nouns at first (1;3–1;7) are exclusively bare (100%) and then predominantly so (92% bare at 1;8–1;9, 63% bare at 1;10–1;11), in striking contrast to the adults.

It is unclear how the heavy preponderance of bare nouns, i.e. the omission of articles, in early child speech could be captured by a frequency-based approach to language acquisition. Unless one can explain why children seem to pay attention to some input frequencies (e.g. definite vs. indefinite articles), but are willing to ignore input frequency patterns in others (e.g. determined nominals being so much more frequent than bare ones), I doubt whether speaking of an input frequency “effect” in the former case explains very much. One might, of course, say that the child only produces bare nouns at first because s/he has to start somewhere and will start with bare
nouns before going on to learn morphological article markers. Note, however, that such a view has little to do with constructivist frequency-driven approaches (e.g. Tomasello 2003; Theakston et al. 2004), according to which the child starts to produce unanalyzed strings that are frequent in the input. As determined nouns are much more frequent in the recorded input than bare nouns (Figure 6), Markus should start producing unanalyzed \([N+Art]\) and \([Art+N]\) chunks. The fact that he exclusively produces bare nouns in the early recordings suggests that the unanalyzed learned string approach is on the wrong track.

5. Potential frequency effect 2: Definite enclitic vs. definite pronominal

Are definite pronominal articles produced rarely and late by Embla and Markus because they are the least frequent article forms in the input? There is indeed a close frequency match between children and adults (recall Figures 4, 5, 6): only 9% of the definite nominals produced by Embla’s caregivers have a pronominal definite article, and 11% of Embla’s, cf. Table 3. 15% of Markus’ caregivers’ definite nominals involve a pronominal article, and 22% of Markus’, cf. Table 4. Thus, the overwhelming majority of definite articles are not pronominal but enclitic.

*Table 3. Definite articles, enclitic vs. pronominal: Embla*

<table>
<thead>
<tr>
<th></th>
<th>Embla’s caregivers (first 5 samples)</th>
<th>Embla (1;8,2–2;1,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite enclitic only</td>
<td>91% (61/67)</td>
<td>89% (143/161)</td>
</tr>
<tr>
<td>Definite pronominal + enclitic</td>
<td>9% (6/67)</td>
<td>10% (16/161)</td>
</tr>
<tr>
<td>Definite pronominal only</td>
<td>0% (0/67)</td>
<td>1% (2/161)</td>
</tr>
</tbody>
</table>

*Table 4. Definite articles, enclitic vs. pronominal: Markus*

<table>
<thead>
<tr>
<th></th>
<th>Markus’ caregivers (all samples)</th>
<th>Markus (1;3,19–2;0,2, all samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Definite enclitic only</td>
<td>86% (540/631)</td>
<td>78% (200/255)</td>
</tr>
<tr>
<td>Definite pronominal + enclitic</td>
<td>12% (75/631)</td>
<td>14% (35/255)</td>
</tr>
<tr>
<td>Definite pronominal only</td>
<td>3% (16/631)</td>
<td>8% (20/255)</td>
</tr>
</tbody>
</table>
Low numbers of prenominal articles in Swedish child speech have sometimes been interpreted as non-targetlike article omissions (e.g. Santelmann 1998: 653–657). As I have argued in Bohnacker (2003: 232–236) and will do so again here, such an interpretation cannot be upheld in light of the input these children receive. Firstly, the input frequency of prenominal definite articles (as documented in the recordings) is very low (see Table 3 and Table 4). This is not very surprising because prenominal definite articles run a much lower chance to occur than enclitic articles in Swedish, simply because prenominal definite articles can only occur in nominal phrases that involve a prenominal demonstrative or an attributive adjective, and attributive adjectives are rare in adult-to-child speech cross-linguistically. (In Markus’ input samples, only 7% of all nominals with a common noun contain an attributive adjective.) Secondly, as was discussed in Section 2, the input – and colloquial Swedish varieties in general – are characterized by variability, i.e. “omissions” of prenominal definite articles from nominals with adjectives.

Let’s take a closer look at Markus’ caregivers. Their 75 doubly determined nominals (definite prenominal + enclitic) break down into three types: 23 with a stressed demonstrative prenominal followed by N+Art as in (7a); 30 with an unstressed prenominal followed by a stressed demonstrative här/där plus N+Art as in (7b), and 22 cases with an unstressed prenominal followed by an adjective plus N+Art as in (7c). In addition, the caregivers produce 17 singly determined nominals with adjectives (Adj+N+Art), e.g. (8).

(7) a. nä inte DEN stol-en Markus.  
   no not the stressed chair-the Markus
   ‘No, not THAT chair, Markus.’ (caregiver, 1;3,19)

b. mamma har den HÄR tröja-n.  
   mum has the here stressed jumper-the
   ‘Mum’s got THIS jumper.’ (caregiver, 1;8,05)

c. var e den lilla flicka-n?  
   where is the little girl-the
   ‘Where’s the little girl?’ (caregiver, 1;3;19)

(8) å så kommer _ lilla hund-en.  
   and so comes little dog-the
   ‘And then comes the little dog.’ (caregiver, 1;4,27)
This means that 44% (17/39) of the caregivers’ definite nominals with an attributive adjective are singly determined, i.e. in about half of the cases the prenominal article is “omitted”, whereas definite nominals with a demonstrative are doubly determined all the time (prenominal + enclitic). This is what the immediate target for the child looks like, and the child productions are quite similar: Markus omits definite prenominal articles from nominals with an adjective, but rarely from nominals with a demonstrative, as shown in Table 5 (raw figures, article omissions are indicated by strike-through).

Table 5. Definite prenominal article: Markus

<table>
<thead>
<tr>
<th>Description</th>
<th>Markus’ caregivers</th>
<th>Markus (1;3,19–2;0,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>demonstrative Det + N+Art</td>
<td>23</td>
<td>35</td>
</tr>
<tr>
<td>demonstrative Det + N+Art</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Art + demonstr. {här/där} + N+Art</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Art + demonstr. {här/där} + N+Art</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Art + Adj + N+Art</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>Art + Adj + N+Art</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>Art + nominalized Adj, no N</td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>Art + nominalized Adj, no N</td>
<td>0</td>
<td>6</td>
</tr>
</tbody>
</table>

What Markus and Embla appear to be doing, then, is presumably to pay attention to input frequencies of definite prenominal and definite enclitic articles, and most certainly to pay attention to input characteristics, that is to the grammar of the immediate target language.

6. Potential input frequency effect 3: Differences in bare noun frequency across languages

Let us now return to articles and article omissions in general (Sections 2 and 3). Could the extent to which children omit articles be correlated with the relative frequency of articleless (bare) nouns in the input? Kupisch (2003, 2004, this volume), for instance, sets out to demonstrate that children start to produce articles later and omit articles for a longer period if they are exposed to a Germanic language with a high frequency of bare nouns. Kupisch (2004: 22–30) convincingly shows that bare nouns in child-
directed adult German are substantially more frequent (18% bare nouns) than in child-directed adult French and Italian, but also that there are frequency differences between the Romance languages (French (6%), Italian (11%)). Kupisch argues that these adult bare noun frequencies affect the children’s rate of acquisition, as most of her German children start producing articles later than the French and Italian children do (Kupisch 2003).

Her proposed bare noun input frequency effect receives support from a cross-sectional elicitation study: German children that are older than their French and Italian peers nevertheless show a higher rate of omission of determiners, with significant differences between the three languages for the youngest children: 59% bare nouns in child German, 41% in child Italian, 19% in child French. (Kupisch 2004: 94–95).

Such a frequency account of bare vs. determined nouns rests on the assumption that children indeed count grammatical categories and have at least a rough idea of what a determiner is.

Can Kupisch’s bare noun input frequency account be extended to Swedish? Like German, Swedish is a Germanic language and has bare nouns. And due to the fact that Swedish allows bare singular count nouns in a variety of contexts (cf. Section 2), the input frequency of bare nouns may be quite high, even though exact figures have not been established for any corpus yet. Kupisch would thus predict that Swedish articles emerge cross-linguistically late. I will show that Swedish child-directed speech has indeed a high ratio of bare nouns, at least as far as the recordings of Embla’s and Markus’ caregivers are concerned, which I hand-searched. To make the database comparable to Kupisch’s (2004), I counted as bare nouns (BNs) bare singular count nouns, bare mass nouns, bare abstract nouns and bare plurals. Proper nouns, (namelike) kinship nouns and pronouns were excluded. Bare noun ratio was established as the number of bare nouns divided by the sum of bare and determined nouns. As shown in Table 6 and Table 7, there are 20% bare nouns in Embla’s caregivers’ and 22% bare nouns in Markus’ caregivers’ speech.

Table 6. Bare nouns: Embla

<table>
<thead>
<tr>
<th>Embla’s parents/caregivers (first 5 samples)</th>
<th>Embla (first 5 samples, 1;8,2–1;10,2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>20% bare nouns (23/116)</td>
<td>52% bare nouns (62/119)</td>
</tr>
</tbody>
</table>
Table 7. Bare nouns: Markus

<table>
<thead>
<tr>
<th>Markus’ caregivers (all samples)</th>
<th>Markus (all samples, 1;3,19–2;0,25)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22% bare nouns (327/1487)</td>
<td>58% bare nouns (567/972)</td>
</tr>
</tbody>
</table>

These Swedish BN percentages should advantageously be verified for a larger caregiver corpus, but they are higher than what Kupisch found for adult-to-child speech in French (6%), Italian (11%) and German (18%). Yet despite the 20%–22% bare nouns in the sampled input, Embla and Markus do not produce articles as late as Kupisch’s German children. Rather, as we have seen, Embla and Markus produce articles early on and reach adult provision levels by 2;0 (Figures 1 and 2). Hence, the input frequency of bare nouns (in Swedish) does not appear to affect the course of acquisition as directly as Kupisch (2004) suggests.

7. Potential input frequency effect 4: Bare noun (subtype) frequency over time

There could be another, language-internal, input frequency effect to do with bare nouns, for instance, a substantial shift in bare noun frequency over time, correlating with child production. There might also be a shift in bare noun composition over time. For instance, a breakdown into four semantic types – bare singular count (e.g. pāse ‘bag’), mass (e.g. ost ‘cheese’), abstract (e.g. sömn ‘sleep’) and plural (e.g. hästar ‘horses’) – in Table 8 reveals that a full 30% of the adults’ BNs are bare singulars. Even though linguists disagree about the acceptability of bare nouns and tend to judge such bare singulars as ungrammatical, we are interested here in what Swedish-speaking parents actually say to their children.

Table 8. Bare nouns by semantic type: Markus’ caregivers (all samples)

<table>
<thead>
<tr>
<th>Semantic type</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bare singular count noun</td>
<td>30% (98/327)</td>
</tr>
<tr>
<td>Bare mass noun</td>
<td>31% (102/327)</td>
</tr>
<tr>
<td>Bare abstract noun</td>
<td>5% (17/327)</td>
</tr>
<tr>
<td>Bare plural noun</td>
<td>34% (110/327)</td>
</tr>
</tbody>
</table>
Table 9 provides a rough breakdown into five distributional syntactic contexts, BNs in isolation (i.e. BNs that constitute utterances of their own, not being part of a sentence), BNs with a copula verb, BNs as the subject or object of a thematic verb, BNs as the object of a preposition, and “other”.

Table 9. Bare nouns by syntactic context: Markus’ caregivers (all samples)

<table>
<thead>
<tr>
<th>Syntactic context</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>In isolation</td>
<td>22% (71/327)</td>
</tr>
<tr>
<td>With copula</td>
<td>33% (109/327)</td>
</tr>
<tr>
<td>With thematic verb</td>
<td>33% (107/327)</td>
</tr>
<tr>
<td>With preposition</td>
<td>11% (37/327)</td>
</tr>
<tr>
<td>Other</td>
<td>1% (3/327)</td>
</tr>
</tbody>
</table>

A breakdown of Markus’ caregivers’ BNs over time shows that there is a decrease in the overall frequency of bare nouns (Figure 7). For ease of exposition, I have grouped the adult productions into four periods, the same periods that Markus’ productions were grouped into earlier (1;3,19–1;7,25; 1;8,5–1;9,7; 1;10–1;11,12; 1;11,25–2;0,25).

Figure 7. Bare nouns produced by Markus’ adult caregivers by stage, percent. Bare nouns out of all nominals (determined and bare, but excluding proper nouns, namelike kinship nouns and pronouns).
The grey columns in Figure 7 show that the adults produce a higher percentage of bare nouns in the early recordings than they do later on: 31% BNs at 1;3–1;7 (92/300), vs. 19%–20% BNs later (1;8–1;9: 19% (74/388); 1;10–1;11,12: 20% (79/398); 1;11,25–2;05: 20% (82/401)). One may be tempted to conclude that there could be a link between the high percentage of BNs in the early input and Markus’ exclusive use of BNs in his early productions (recall Section 2, Figure 3).

There is also a shift in the semantic distribution of BNs (cf. Table 8) over time: in the early input samples, bare singular count nouns predominate (43%), but become less frequent later on. This development is sketched by the black line in Figure 7. At 1;3–1;7, there are 43% (40/92) bare singulars out of all BNs, corresponding to 13% (40/300) bare singulars out of all nominals, versus 3% (12/388) bare singulars at 1;8–1;10, 6% (22/398) at 1;10–1;11,25, and 6% (24/401) at 1;11,25–2;0,25.

Also noteworthy is a shift in the syntactic distribution of BNs: In the early samples, Markus’ caregivers frequently produce bare nouns in isolation (1;3–1;7: 33% of all BNs (30/92)), but do so much less frequently in the later recordings (1;8–1;9: 22% (16/74); 1;10–1;11,12: 20% (16/79); 1;11,25–2;0,25: 11% (9/82)). This development is indicated by the white columns in Figure 7. Caregivers use such BNs in isolation for instance when teaching the child a noun, naming, and confirming what the child has said (e.g. näsa. (‘nose’); ost. (‘cheese’); ja ljus ja. (‘yes candle yes’); pojke (‘boy’); kaffe ja. (‘coffee yes’); å öga ja. (‘and eye yes’).

Another shift concerns BNs in copula constructions which are frequent at first (1;3–1;7: 38% (35/92)), but less frequent later (1;11,25–2;0,25: 17% (14/82)). A reverse shift is found for BNs in thematic verb constructions. BNs as the subject or object of thematic verbs are infrequent in the early input samples (1;3–1;7: 16% (15/92) but make up the majority of BNs in the late samples (1;11,25–2;0,25: 62% (51/82)).

For Markus’ caregivers then, we can document a decrease in overall bare noun frequency and a shift in bare noun composition over time. BNs in isolation are frequent in the earliest samples, but are largely supplanted by nouns in other syntactic contexts. Semantically, bare singular count nouns are frequent in the earliest samples, but later become rare. This suggests that very early Swedish child-directed speech has a larger proportion of just the kind of bare nouns that the child produces most: bare singular count nouns and bare nouns in isolation. Again, these findings for Markus and his caregivers would need to be tested for other Swedish adult-to-child corpora, but they could point to a link between the use of bare nouns in the
The role of input frequency in article acquisition in early child Swedish

early input and Markus’ exclusive use of BNs in his early productions. However, a chicken-and-egg problem arises. Are Markus’ BNs the result of his following an input pattern, or do the caregivers in the early samples produce more BNs, more bare singular count nouns, and more BNs in isolation because they are adapting to the child and his use of BNs? Without recordings of the same adults in other situations, this question is impossible to answer.

Is there an input frequency effect, then, with regard to bare nouns? I also find this question difficult to answer. Bare nouns (and certain subtypes in particular) in the recorded input are frequent (22%), and in the earliest recordings especially so (31%). Markus produces bare nouns at an extremely high frequency, at first exclusively so (recall Figure 3): 100% BNs at 1;3–1;7, 92% BNs at 1;8–1;9. But these figures also indicate a large discrepancy between caregivers and child, which, I should think, makes it impossible to explain the child’s acquisitional path as an effect of input frequency.

However, investigating caregivers’ article and BN (subtype) frequencies does help us discover that bare nouns and bare singular count nouns are more frequent in Swedish and occur in more contexts than previously assumed.

8. Conclusion

In this paper, I have investigated the frequencies of certain morphological elements (articles and article types) and their absence (bare nouns) in Swedish child and adult-to-child speech. This has proved to be a worthwhile enterprise, as the adults use nominals somewhat differently from what the average linguist may think. Of course, I have only looked at two children (Embla and Markus) and their caregivers, and studies of additional corpora would be well in place. Also, findings are tentative, as the recordings only represent a fraction of the actual input these children receive. Nevertheless, indefinite articles in the input are slightly less frequent (45%) than definite ones (55%), and doubly definite nominal phrases (12%) are vanishingly rare compared to singly definite ones (88%). Prenominal definite articles are rare and often “omitted” from nominals with attributive adjectives, much more often so than what is assumed in grammar books and theoretical accounts of Swedish. And bare nouns, including bare singular count nouns, are also frequent and occur more frequently and in more
contexts than reference grammars and non-corpus-driven research suggest. Such results should be of interest to any Swedish grammarian and theoretical linguist, and for the acquisitionist they provide a better yardstick to measure early child language productions against. Moreover, investigating the different kinds of nominals in caregiver speech and their frequency patterns enables us to make interesting new cross-linguistic comparisons of adult-to-child speech, as Kupisch’s (2004, this volume) results for bare nouns in German, Italian and French and my results for Swedish have shown.

When it comes to input frequency as an explanatory tool for the acquisition of articles, however, I am less enthusiastic and much more doubtful. What does “finding an input frequency effect” actually mean? Does it simply mean finding similar frequencies regarding a particular morphological marker (here, a particular article)? In this case, there are input frequency “effects”, since there are close matches: as shown in Sections 3–7, Embla and Markus – like the adults – produce higher percentages of definite articles than indefinites, and much higher percentages of definite enclitic articles than definite prenominal ones. Interesting as such matches may be, I have trouble understanding what the actual “effect” of input frequency would be here. To my mind, finding a match does not explain the process of acquisition, we will still need a theory of language acquisition – of whichever theoretical persuasion – to explain that. And let’s not forget that we also find severe discrepancies, or mismatches, regarding articles in Swedish caregiver and child production. This is especially evident for bare nouns (Sections 6–7). Even though the same types of nominal phrases occur in both children and adults, Embla (1;8–2;1) and Markus (1;3–2;0) produce many more bare nouns than the adults do, and Markus, for whom we have earlier recordings, produces only bare nouns at first (100% at 1;3–1;7; 92% at 1;8–1;9) and then still predominantly bare nouns (1;10–1;11, 63%), in marked contrast to the adults.

Or does “finding an input frequency effect” mean that the child first reaches target criterion for the morphological marker that is most frequent in the input, and last for the one that is least frequent? I do not think that my results are particularly telling in this regard. Definite enclitic and indefinite prenominal articles reach target provision more or less at the same time (2;0), while the raw figures for definite prenominal articles are very low, with the added complication of unstressed prenominal definites being optional for the adults. This makes it difficult to establish obligatory contexts and a target provision criterion (Section 5).
A third possibility regarding input frequency effects is that the most frequent morphological marker in the input emerges first in child language production, and the least frequent one last. On this interpretation, we find a clear input frequency effect for the definite enclitic vs. the definite prenominal article. The former is highly frequent in the recorded input and is the article to emerge first (in Markus), as well as being the first article to occur with any regularity in Markus and Embla. By contrast, the definite prenominal article is superbly rare in the input samples, emerges later than the other articles in Markus and Embla, and remains a rare occurrence in both children. An additional frequency effect may be argued to be found for Embla, where indefinite articles are much rarer than definite ones in the input samples, and in the child do not occur with any regularity until late. However, Embla’s corpus is small and therefore less reliable, and Markus’ larger corpus disturbs the pretty picture: while indefinite articles in Markus’ productions do emerge later and continue to be much less frequent than definite articles (as for Embla), in his case, this cannot be attributed to low-frequency input, because indefinite articles in his input are very common indeed (45% of all articles). Thus, for several months, Markus produces bare nouns instead of nominals with an indefinite article, despite the high input percentage of indefinite articles.

It thus seems to me that effects of input frequency are hard to show and even harder to prove. To my mind, the frequencies of certain morphological markers (here, articles) and the frequency of the absence of such markers (here, bare nouns) in the input are simply reflexes of the characteristics of the immediate target language, i.e. epiphenomenal reflexes of the grammar to which the child is exposed. At the very beginning, the child’s system – at least in production – is impoverished in many ways (e.g. no articles), while later it has become much closer to the language system of the adults, and one way of showing this are the close caregiver-child frequency matches found in the later recordings of Embla and Markus at 2;0. But documenting mismatches and matches of input and output is not enough to explain acquisition. Perhaps we are better advised to look beyond input frequencies for explanations for the initial absence of articles, for the omission of mainly indefinite articles, and for the rapid – and cross-linguistically early – acquisition of articles in child Swedish. In pursuing this goal, we might, for instance, investigate the morphophonological form of articles and the child’s development of segmental and suprasegmental phonology, including metrical structure, as well as situational context and pragmatics (e.g. Bohnacker 2003: 225–247; Kupisch 2004; Lleó 1998).
Notes

1. The prenominal definite article also occurs with nominalized adjectives, i.e. in definite nominals without a noun, e.g. det svarta _ (the black ‘the black one’). Possessive determiners (e.g. mitt ‘my’) are also prenominal (e.g. mitt får ‘my sheep’), but will not be discussed here.

2. A similar distribution appears to hold for Danish and Norwegian bare singulars (cf., e.g., Hansen 1998: 42–46; Borthen 2003).

3. A note on age and MLU: In what follows, I have chosen to only give the child’s age, even though I am aware that age-matching can be problematic. I do not think that MLU-matching would be a better criterion to measure language proficiency, however, firstly because it can skew cross-linguistic comparisons, and secondly, because MLU is not an independent measure when it comes to articles: When a child begins to produce articles, that in itself increases her MLU. Ideally, one might want to combine several criteria such as age, vocabulary size, increase in word types, total number of utterances, upper bound of utterances, etc.

4. Even though articles are often omitted, articles do not seem to be misapplied. When hand-searching the transcripts, I found only one misuse, namely Markus’ out-of-the-blue or egocentric use of a definite article. As illustrated in (4) above, Markus says andra gabbe-n ‘the other man’, to which his father reacts uncomprehendingly: ‘The other man? But there is only one man in the car. There should be another man?’.

5. In a related vein, Kupisch (2004) found that definite articles in adult-to-child speech were more frequent in Italian (70% of all DPs) than in French (49%) and German (51%), and that definite articles emerged earlier in child Italian than in child German.

6. This strong asymmetry in Embla’s caregivers is likely to be a sampling artefact (short samples in particular situational contexts), and not attested for other, larger corpora, such as Markus’ caregivers.

7. This includes all definite enclitic articles, also those in double determination constructions.

8. The adjectives in Markus’ caregivers’ doubly determined and singly determined nominals are restricted to a handful of types. lilla (‘little’) and andra (‘other’) predominate.

9. For Embla, the amount of transcribed input available is small, with a total of 6 prenominal definite articles, which makes input frequency investigations unfeasible (3 stressed demonstrative Det + N+Art, 2 unstressed Art + demonstrative här/där + N+Art, 1 unstressed Art + Adj + N+Art, plus 1 prenominal article.
omission with a nominalized Adj). Embla herself produces 15 stressed demonstrative Det + N+Art, 1 unstressed Art + demonstrative dår + N+Art, 1 unstressed Art + Adj + N+Art, and 8 singly determined nominals with attributive adjective, i.e. 8 prenominal article “omissions”.

10. Consistency may play a role here too. If articles are frequent (i.e. BNs are rare) and consistently provided in the input (i.e. BNs are licit in only few (syntactic/semantic) contexts, as is the case in French), then article acquisition will be facilitated (Kupisch 2003, 2004).

11. A similar chicken-and-egg problem concerns the use of proper nouns (e.g. Markus, mamma ‘mom’) with first and second person reference, instead of first and second person pronouns (e.g. jag, min, du, din ‘I, my, you, your’). Such nominal style is not used in Swedish adult-to-adult speech, but occurs both in Markus’ speech and child-addressed caregivers’ speech. Yet, who is imitating whom here? For discussion, see Bohnacker (2003: 202–209).

References

Anderssen, Merete
2006 The acquisition of compositional definiteness in Norwegian. Ph.D. diss., Tromso: Department of Language and Linguistics, University of Tromso.

Bley-Vroman, Robert

Bohnacker, Ute

Borhen, Kaja

Bottari, Piero, Paola Cipriani & Anna Maria Chilosi
Clahsen, Harald, Sonja Eisenbeiß & Anne Vainikka

Crisma, Paola & Elisabetta Tomasutti

Eisenbeiß, Sonja

Emslie, Hazel C. & Rosemary J. Stevenson

Hansen, Gyde

Hyams, Nina

Jackendoff, Ray

Kupisch, Tanja

this vol. Testing the effects of frequency on the rate of learning: Determiner use in early French, German and Italian.
Langacker, Ronald W.

Lange, Sven & Kenneth Larsson

Lange, Sven & Kenneth Larsson

Lieven, Elena, Heike Behrens, Jennifer Speares & Michael Tomasello

Lleó, Conxita


Lleó, Conxita & Katherine Demuth

Radford, Andrew


Roeper, Thomas

this vol. What frequency can do and what it can’t.

Santelmann, Lynn
1998 The acquisition of definite determiners in child Swedish: Metrical and discourse influences on functional morphology. In BUCLD 22
Söderbergh, Ragnhild

Strömqvist, Sven, Ulla Richthoff & Anders-Börje Andersson

Svartholm, Kristina

Theakston, Anna L., Elena V. M. Lieven, Julian M. Pine & Caroline F. Rowland

Tomasello, Michael

Wexler, Ken

Yang, Charles