Chapter 19. Article Acquisition in English, German, Norwegian, and Swedish

Published by

Leow, Ronald P., et al.
Little Words: Their History, Phonology, Syntax, Semantics, Pragmatics, and Acquisition.

⇒ For additional information about this book
https://muse.jhu.edu/book/13060
Article Acquisition in English, German, Norwegian, and Swedish

TANJA KUPISCH, MERETE ANDERSSEN, UTE BOHNACKER, AND NEIL SNAPE
University of Calgary, University of Tromsø, Uppsala University, and Hokkaido University

ARTICLE OMISSION is a well-documented phenomenon in early child speech. Interestingly, children differ in terms of how extensively they omit articles depending on their age and what language(s) they are exposed to. Different accounts have been proposed to account for this cross-linguistic variation. One of the most widely discussed models is the nominal mapping parameter (NMP), originally proposed in Chierchia (1998), which relates variation in child language to the syntactic and semantic properties of noun phrases across languages (e.g., Chierchia, Guasti, and Gualmini 1999; Guasti and Gavarró 2003; Guasti et al. 2004). Other influential accounts of determiner omission have been formulated in prosody-oriented research (e.g., Gerken 1991, 1994; Lleó 1998, 2001; Lleó and Demuth 1999; Roark and Demuth 2000; Demuth, McCullough, and Adamo 2007). So far, no common agreement has been reached.

This chapter presents a study on article acquisition in English, German, Norwegian, and Swedish, where article use is subject to similar syntactic and semantic conditions. Hence the NMP predicts similar acquisition patterns. In terms of their metrical structure, noun phrases in these languages differ considerably. Accordingly, different acquisition patterns are expected.

The chapter is structured as follows: In the first section, we present an overview of articles and article use in the four languages under discussion. In the second section, we introduce the NMP and two prosodic accounts. The third section presents our results, discussing them in light of the aforementioned models. Our data suggest that the NMP should be discarded as empirically false, while supporting prosodic approaches to article acquisition. Nevertheless, as we conclude in the fourth section, even prosodic accounts do not provide an all-encompassing explanation for children’s omission of articles.

Article acquisition in these four languages has been studied in previous work (for monolingual first language [L1] acquisition, see, e.g., Brown 1973; Radford 1990; Abu-Akel and Bailey 2000; Demuth, McCullough, and Adamo 2007, for English; Clahsen, Eisenbeiss, and Penke 1996; Penner and Weissenborn 1996; Lleó 1998, 2001; Eisenbeiss 2002; Kupisch 2006, 2007, for German; Anderssen 2005, for Norwegian;
Bohnacker 1997, 2004, 2007; Plunkett and Strömqvist 1992; Santelmann 1998; Svartholm 1978, for Swedish). However, no study has provided a systematic comparison of these languages.

Articles in English, German, Norwegian, and Swedish

The four languages all have definite and indefinite articles. Indefinite articles occur prenominally. Definite articles are prenominal in English and German but postnominal in Norwegian and Swedish. [The articles in (1) are marked by italics.]

(1) En. *a house vs. the house Ge. *ein Haus vs. das Haus
No. *et hus vs. hus-e Sw. *ett hus vs. hus-et

Prosodically, articles in these languages differ in terms of whether they constitute clitics or feet. English articles represent free clitics (2a) (Selkirk 1996). In the two Scandinavian languages, the prenominal indefinite article is proclitic to the noun, while the postnominal definite article is enclitic. Traditionally, the suffixal article in Norwegian (and Swedish) is implicitly taken to be of the kind represented in (2b). However, recent research suggests that the suffixal article is an affixal clitic (2c) based on the fact that the addition of the suffixal article does not alter the pitch accent of monosyllabic nouns (Morén 2007). German articles may be reduced or unreduced. Nonreduced articles have been analyzed as phonological words on their own, that is, they form separate feet (2d) (Wiese 1996, p.c.). Reduced articles may be enclitic, being subsumed under the host, which can be a preceding verb (2b) or a preposition (2d). In sentence-initial position, reduced articles may be proclitic (2a).1

(2) a. free clitic b. internal clitic c. affixal clitic d. prosodic word

\[
\begin{array}{cccc}
PP & & PP & PP \\
\hline \\
\text{fnc} & PW & PW & PW \\
\hline \\
\text{lex} & \text{fnc} & PW & \text{fnc} \\
\end{array}
\]

The four languages are largely similar in terms of the syntactic, semantic, and pragmatic conditions of article use. Generally articles are obligatory with singular count nouns, regardless of whether the reading is specific (as with the verb see) or generic (as with the verb like). The absence of an article (or any other determiner) results in ungrammaticality [cf. (3)].2

(3) En. *I see/like _ cat. Ge. *Ich sehe/mag _ Katze.

If the noun is plural or mass, it may appear bare. In this case, the noun phrase (NP) has a nonspecific or a generic reading, depending on the verb [cf. (4)]. Definite articles are allowed in these contexts, but they render the NP specific.
Besides these parallels, there are also some interesting differences. In fact, both Scandinavian languages sometimes allow bare singular count nouns whose meaning is slightly different from the corresponding indefinite marked nouns. The translation equivalents in English and German are ungrammatical. This point is usually not addressed in the linguistics literature and, despite its relevance, not mentioned in any papers dealing with the NMP (cf. Bohnacker 2004, 2007, 53–54, for details).

Based on these observations, one may wonder whether it is still valid to test the NMP comparing these languages. We think it is, because according to Chierchia (1998, 356–57, 400–401) the Germanic languages pertain to one and the same parameter setting.

The Nominal Mapping Parameter (NMP)
The NMP (Chierchia 1998) is concerned with the question of how semantic types of NPs are mapped onto syntax. Chierchia proposes that languages can be subdivided into three types according to what their nouns denote. Each of these types is representative of a parametric setting: In the Chinese-type language, all nouns are typically masslike. They come out of the lexicon as arguments and can be directly mapped onto syntax without projecting a determiner phrase (DP). In this type of language, nouns are always bare and there is no plural morphology. The Romance-type language represents the opposite case. Here, nouns are predicates, and D must be projected to convert them into arguments, which can then appear in argument positions. These languages have plural morphology and nouns must not be bare. In the third type of language, the Germanic type, languages can be either predicates or arguments, depending on whether they are count or mass. Only in the former case must D be projected. Nominals in all languages can be type-shifted (from predicates to arguments, and vice versa; see Chierchia 1998). This is necessary because nouns that are inherently count can sometimes be used as if they were mass, and vice versa.

Chierchia, Guasti, and Gualmini (1999) outline the predictions of the NMP for acquisition as follows: When acquiring articles, children pass through predetermined stages corresponding to parameter settings in the following way. The Chinese setting represents the default setting (stage 1). During this stage, children consistently omit determiners as if their target language were an articleless language. If this setting does
not correspond to the target language, the discovery of plural morphology triggers resetting of the parameter to the Germanic setting where bare nouns and determiner-noun sequences appear to be in “free variation.” During this stage, children have to figure out which nouns are count and which are mass, and article omissions in obligatory contexts result from the misclassification of (particular) count nouns as mass. Children exposed to a Romance-type language have to reset the parameter again. They can do so very quickly on the basis of positive evidence, that is, when discovering that nouns are consistently used with articles (rather than varying with bare nouns). The learning task is assumed to be more time-consuming in the Germanic-type languages, that is, it results in a more extended period of bare-noun use, because Germanic-learning children have to figure out for each noun separately whether it is mass or count. In Chierchia’s typology, the Germanic languages are associated with one and the same parameter setting. Hence the NMP predicts that children acquiring English, German, Norwegian, and Swedish show similar patterns of article use and omission.

**Prosodic Accounts**

Prosody-oriented research has resulted in several different accounts on the acquisition of articles. We discuss two of them, which we refer to as *trochaic templates* and *bootstrapping via lexical models*.

**Trochaic templates.** It is generally assumed that trochaic patterns, that is, words or phrases consisting of a strong syllable followed by a weak one (SW), are unmarked as compared with iambic patterns, that is, words or phrases consisting of a weak followed by a strong syllable (WS). Experimental research has shown that nine-month-old American infants listen longer to lists of items that conform to the predominant strong-weak stress pattern of English than to lists that do not display this pattern (Jusczyk, Cutler, and Redanz 1993). Furthermore it has long been observed that children are less likely to preserve the initial syllable in the pronunciation of words like *baNAna* or *giRAFfe* than the final syllable of *CANdy* or *DONkey*, which also suggests that English children pay attention to SW structures. Gerken (1991) argues that the omission of various function words, including articles, should be explained in terms of a dispreference for iambic structures, as these elements are often prosodified as the pretonic syllables of an iamb. She provides empirical support from an imitation task: Children had to imitate utterances with weakly stressed syllables, including pronouns and articles. The children omitted extrametrical syllables from iambic structures, which led Gerken to suggest that child utterances have to fit what she referred to as a Trochaic Template.

English and most German NPs consisting of a determiner and a noun do not fit the trochaic template because the unstressed article precedes the noun, as in En. *a/the house* (WS), Ge. *(ei)n/das Haus* (WS) (and the majority of nouns are mono- or bi-syllabic). The same is true for Norwegian and Swedish indefinite marked NPs, for example, No. *et hus* (WS) and Sw. *ett hus* (WS). By contrast, most Norwegian and Swedish NPs with a suffixed definite article fit the trochaic pattern, for example, No. *hus-e, katt-a* (SW) and Sw. *hus-et, katt-en* (SW).
If it is true that children show a preference for trochaic templates in acquisition, we would expect to see different acquisition patterns in the children acquiring English and German on the one hand, and children acquiring the Scandinavian languages on the other hand. More specifically, two predictions may be formulated: First, if children produce unstressed syllables more often in trochaic patterns than in iambic patterns, Swedish- and Norwegian-learning children should produce more articles than English- and German-learning children. Second, there should be an asymmetry in the Scandinavian languages between definite articles, which fit the trochaic patterns, and indefinite articles, which do not.

Prosodic bootstrapping via lexical models. According to Lleó and Demuth (1999), children’s use of articles is bootstrapped through the presence of lexemes exhibiting an SWS structure in the input. The authors compare Spanish-learning to German-learning children. As Spanish contains many trisyllabic WSW nouns, Spanish-speaking children are frequently exposed to WSW structures at the lexical level, that is, structures that are also required in the production of articles. German lexemes, by contrast, are mostly mono- or disyllabic, with the initial syllable being strong. The model is couched in terms of optimality theory, where WSW structures are assumed to require the violation of the Exhaustivity Constraint because they involve the production of an extrametrical syllable, that is, a syllable not immediately governed by a foot. Lleó and Demuth assume that because Spanish children have more evidence for the violation of Exhaustivity on the lexical level, they discover sooner than German children that violating Exhaustivity is a requirement in the production of utterances in their target language. Hence they produce articles earlier. This approach goes one step farther than the Trochaic Templates approach by motivating the early acquisition of articles in Spanish as opposed to German.

In the four languages we examine, the great majority of words have stressed initial syllables. An analysis of root nouns in Norwegian (facilitated by the Text Laboratory at Oslo University) indicated that only 648 (4.7%) of 13,848 non-derived nouns have initial weak syllables, the vast majority being loan words unlikely to occur in the input of small children. Because the four languages are typologically closely related, we do not expect their syllable structure to be noticeably different. Hence this model predicts that prenominal articles emerge late in all four languages, while definite articles should occur early in Norwegian and Swedish because they correspond to the predominant metric pattern in the target language.

The Study
Our analysis is based on longitudinal and cross-sectional data. For English and German, both cross-sectional and longitudinal data have been used. The English data represent different children at the ages of 1;10, 2;0, 2;1, and 2;5. For German, two children were examined longitudinally (1;6–2;5 and 1;8–3;0), and ten more children were studied at 1;10, 2;1, and 2;5. The English and German data were taken from the Manchester corpus and from the Szagun corpus, respectively, both available through CHILDES (Theakston et al. 2001; Szagun 2001; McWhinney and Snow 1990). For Norwegian and Swedish, where fewer data were available, we used only
longitudinal corpora. The Norwegian data were collected by Merete Anderssen (Anderssen 2005). The Swedish corpus Markus was collected by Sven Strömqvist (Plunkett and Strömqvist 1992) and is available through CHILDES. The corpus Embla is from the Stockholm-based Swedish Child Language Syntax Project (Lange and Larsson 1973).

We compared the children in terms of mean length of utterances (MLU) rather than age to make our analysis comparable to others, particularly Chierchia et al. (1999). Our MLU is based on words rather than morphemes because the languages differ noticeably in the amount of bound morphology. An overview of the files analyzed is presented in Table 19.1. We calculated the percentage of article use by determining the number of contexts in which an article was used from the total of contexts in which native speakers would use an article. We focus on the production of articles (rather than the whole class of determiners) to render our analysis comparable to previous ones testing the previously mentioned models. We ignored contexts in which bare nouns were used correctly, such as (4) through (5).

Figure 19.1 provides an overview of article omission in all four languages. To be able to run statistics, we also subdivided the data into MLU stages (MLU 1–1.49, MLU 1.5–1.99, etc.; figure 19.2). Between MLU 1 and MLU 2.5, there are statisti-

### Table 19.1

<table>
<thead>
<tr>
<th>MLU</th>
<th>English</th>
<th>German</th>
<th>Norwegian</th>
<th>Swedish</th>
</tr>
</thead>
<tbody>
<tr>
<td>1–1.49</td>
<td>7 (265)</td>
<td>26 (1,902)</td>
<td>0</td>
<td>14 (367)</td>
</tr>
<tr>
<td>1.5–1.99</td>
<td>13 (666)</td>
<td>10 (913)</td>
<td>7 (529)</td>
<td>5 (184)</td>
</tr>
<tr>
<td>2.0–2.49</td>
<td>8 (547)</td>
<td>10 (1,106)</td>
<td>3 (262)</td>
<td>3 (164)</td>
</tr>
<tr>
<td>2.5–2.99</td>
<td>7 (408)</td>
<td>6 (708)</td>
<td>4 (318)</td>
<td>3 (210)</td>
</tr>
<tr>
<td>3.0–4.0</td>
<td>3 (280)</td>
<td>3 (424)</td>
<td>13 (1,121)</td>
<td>1 (107)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>35 (2,166)</td>
<td>48 (4,629)</td>
<td>27 (2,230)</td>
<td>26 (1,032)</td>
</tr>
</tbody>
</table>

**Figure 19.1** Article Suppliance in English, German, Norwegian, and Swedish (stages)
cally significant contrasts in the use of articles in obligatory contexts, between English and German on the one hand and Norwegian and Swedish on the other hand. Between MLU 2.5 and MLU 3, the German children catch up with the children acquiring Norwegian and Swedish. Unexpectedly, the English children lag behind the German children here, and so does the Norwegian child with respect to the Swedish children. With an MLU above 3, children in all four languages use articles in more than 80 percent of all obligatory contexts.

Discussion

Our results falsify the NMP, which predicts similar acquisition patterns within Germanic languages. Furthermore, there is evidence against the assumption that article omissions in obligatory contexts in the free variation stage result from a misclassification of particular count nouns as mass, because article omission and realization sometimes occur within the same recording and even in parallel contexts of use. For example, at the age of 2;5.4, the English-speaking child Anne produced both *I want have a drink now* and *Want to have drink again.* At 2;1.21, the German child Martin produced both *Neemann,* “snowman,” and *ein Neemann,* “a snowman,” in two separate utterances, one right after the other. At 1;10.4, the Swedish child Markus produced the *bygga ett tåg,* “build a train,” right after producing *bygga tåg,* “build train.” At 1;10.4, Ina produced *en mann,* “a man,” right after producing *mann,* “man.” (See Cheng and Sybesma 1999; Munn and Schmitt 2001; Kupisch 2006, 103–4, for more counterevidence to the NMP, both theoretical and empirical.)

At the same time, the results lend support to the metrical template approach, which predicts lower rates of article omission in the Scandinavian languages. Moreover, in the Scandinavian languages, prenominal indefinite articles are omitted more than suffixal articles (figure 19.3). Again, this is exactly what the metrical templates approach predicts (see also Santelmann 1998). These results can also be captured by assuming that articles are bootstrapped via lexical models because Norwegian- and Swedish-learning children have lexical models for the metrical structure of nouns with definite articles, while German and English children do not. Figure 19.3 shows that the contrast between English/German and Swedish/Norwegian can largely be

![Figure 19.2 Article Suppliance in English, German, Norwegian, and Swedish](image-url)
attributed to definite article use, while indefinite articles are used to similar extents and from the same age in all languages.

With regard to indefinite article use, it should be noted that although the idea of bootstrapping via lexical models is supported by our results with regard to the definite suffix in Norwegian and Swedish, it raises new issues. In fact, Lleó and Demuth’s (1999) hypothesis that children start to produce articles earlier because their target languages contain many words with unfooted syllables cannot be applied to the languages under discussion here. First, all the children’s utterances we analyzed show no production of any lexemes exhibiting WSW structures—this is not surprising given that the target languages exhibit few such lexemes. For example, none of the lexemes in the first Norwegian file involve an unfooted syllable (see Anderssen 2005, 277), but Ina already produces WSW structures in multiword utterances. The Swedish child Embla produces WSW utterances in the first file at 1;8.2, but she still omits articles in 26 percent of all obligatory contexts. Markus produces some WSW utterances between 1;7 and 1;9, for example, de LAMpa, “it-is lamp,” at 1;9.3, while omitting 100 percent of all indefinite articles. This suggests that WSW lexemes in the input cannot be the factor that catalyzes prenominal article use, as extrametrical syllables occur in multiword utterances before they appear at the word level. In short, the idea of bootstrapping via lexical models predicts the early emergence of definite articles in the Scandinavian languages (many nouns having the metrical structure SW). However, it fails to explain how prenominal articles are acquired in these languages and why extrametrical syllables occur in multiword utterances while indefinite articles are omitted.

**Does Input Frequency Provide an Alternative Solution?**

We mentioned in the first section that indefinite articles appear to be more “optional” in the Scandinavian languages than in English and German. This raises the question whether the higher amount of definite as opposed to indefinite marked NPs in Norwegian and Swedish could result from different distributions in the adult input to the child. Similarly, the higher amount of bare nouns in English and German than in the
Scandinavian languages may result from respective distributions in the input. We conducted two analyses to test these possibilities, examining for each language the amount of definite as compared with indefinite articles and the amount of bare nouns from the total of NPs. Table 19.2 shows that NPs with definite articles are more frequent than NPs with indefinite articles in all four languages. However, only Norwegian and Swedish children show a higher use of definite articles in the onset of article use. Hence input frequency cannot provide an explanation for the early use of definite articles in Norwegian and Swedish.

Table 19.3 shows the amount of bare nouns from the total of NPs. A total of 894 noun phrases were analyzed in English, 643 in German, 552 in Norwegian, and 1,487 in Swedish. Proper names were excluded: Norwegian exhibits the highest amount of bare nouns, although articles in this language are acquired early. English exhibits the lowest amount, although articles in this language are acquired comparatively late. German and Swedish exhibit similar numbers, although children learning these languages show different patterns of article use. Hence input frequencies do not provide an alternative account for our findings.

Conclusions
Our analysis has shown that the NMP cannot account for the distribution of articles in child Germanic. It would only be tenable by adding to the original proposal auxiliary hypotheses to explain the variation attested across Germanic languages. However, auxiliary hypotheses would make the NMP lose much of its predictive force. Hence prosodic models, which can capture this variation without any additional assumptions, are preferable. The metrical template approach and the lexical bootstrapping approach correctly predict the early appearance of articles in the Scandinavian languages and the higher amount of definite as opposed to indefinite articles.

We have further shown that input frequencies do not provide an alternative solution to our findings because the distributions in the input do not match the acquisition patterns, either with regard to the amount of bare nouns found in the child data or with regard to the distribution of definite as opposed to indefinite articles. Overall we agree with Lleó and Demuth (1999), Lleó (2001), and Demuth, McCullough, and...
Adamo (2007) that phonological patterns could be responsible for the variation we observe in the emergence of articles across languages. The fact that definite articles are part of trochees in Norwegian and Swedish seems to render them more salient, which can explain the children’s early awareness of them.

Nevertheless, we admit that phonological approaches cannot provide an all-encompassing explanation of article acquisition. First, neither of the two prosodic models provides an account of why and how children acquire indefinite articles, which do not form trochees with the nouns they accompany and for which the languages provide no prosodic models in the domain of words (in the sense of Lleó and Demuth 1999). Second, neither prosodic model accounts for the (seemingly) optional use of articles at later stages of acquisition (especially MLU 2 to 3). Third, a closer look at the data from a qualitative perspective indicates that the distinction between Scandinavian definite articles (which fit the trochaic template and are acquired early) and indefinite articles (which do not fit the template and are acquired late) is too coarse. It makes a number of predictions that are not borne out: First, it leaves unexplained any omission of the definite article with monosyllabic nouns, for example, No. hus-e, Sw. hus-et, “house-the,” bil-n, Sw. bil-en, “car-the.” Such omissions are rare, but they do occur (see Bohnacker 2004; Anderssen 2005). Second, the trochaic template approach predicts the omission of the third syllable with nouns such as Sw. Album-et, “album-the,” where the definite article adds a syllable to a trochee, as well as the omission of the third syllable in plural nouns, for example, Sw. HÄStar-na, “horses-the.” Such (predicted) omissions are very rare (see Bohnacker 2004, 232–36 for a more detailed discussion). Third, while German children omit articles, they often produce combinations of deictic pronouns and nouns, for example, da Messer, “there knife,” which have the same metrical structure as noun phrases with ( unreduced) articles, for example, das Messer, “the knife.” Hence there are some empirical data that metric approaches cannot capture. These need to be discussed in future research.

NOTES

1. According to Lleó (2001, 33), cliticized articles can never be initial in German. We think that there are exceptions. In High German, sentence-initial indefinite articles are often reduced, typically with topicalized noun phrases (NPs), e.g., N’Haus kann ich jetzt nicht kaufen, “A house I can’t buy now.”

2. We consider a NP to be specific if it refers to one or more particular entities and nonspecific if it does not refer to any particular entity.

3. As an anonymous reviewer pointed out to us, Aoun and Li (2003) show that, contrary to Chierchia’s (1998) claims about Chinese, Mandarin does have a restricted (yet frequent) plural marker –men, which is obligatory on plural personal pronouns and optional on nouns denoting humans. The observation is problematic for the NMP, both because the NMP claims that articleless languages have no plural marking and because plural markings are supposed to trigger parameter resetting, from the Chinese to the Germanic setting.

4. The original order proposed was Chinese → Romance → Germanic. This ordering is problematic because “Romance” is a subset of “Germanic,” if the Romance languages are defined as languages without any bare nouns. All nominals that Germanic-learning children in the Romance setting hear would be consistent with their grammar, and there would be no motivation to reset the parameter. This original proposal was revised in Chierchia, Guasti, and Gualmini (1999). Yet the fact that all Romance languages display some bare nouns raises the issue whether “Romance” and “Germanic” instantiate different parameters.
5. SW refers to the metric pattern, while Sw. refers to the language Swedish.
6. This is a simplification because it assumes that NPs always occur in isolation (which is not true for all cases). Articles may be prosodified with preceding material in the context of larger utterances, for example, if they follow a monosyllabic verb or preposition.
7. However, note that the “input” is determined through the children’s utterances: The proportion of WSW structures is estimated based on the proportion of the children’s vocabulary that involves a WSW structure (in the target language). A slightly different approach has been outlined in Roark and Demuth (2000).
8. We performed chi-square tests for comparing each language contrast in each MLU stage.
9. One reason may be our treatment of bare nouns after prepositions in German. German articles may fuse with preceding prepositions, for example, in den Topf, “into the pot” → in ’Topf. The bare noun structure in Topf only differs by a lack of lengthening on the final nasal of in. Because such acoustic differences are hard to perceive in child speech, we exclude such contexts from our counts, but they may be an area where variation is most persistent.
10. At the last MLU stage (figure 19.3), the Norwegian child Ina uses a surprisingly high number of illegitimate bare nouns as compared with the other children, particularly the two Swedish ones. Most likely this is the result of an individual characteristic of Ina. Studies of other grammatical domains suggest that Ina allows optionality for a prolonged period (Westergaard 2007).

REFERENCES


This page intentionally left blank