

Article Omission: across Child Languages and across Special Registers

Maria Teresa Guasti¹, Joke de Lange², Anna Gavarró³ & Claudia Caprin¹
Università di Milano-Bicocca¹, Universiteit te Utrecht², Universitat Autònoma de
Barcelona³

1. Introduction

Children speaking a wide variety of languages omit articles in their earliest productions, a phenomenon that has been documented over the last five years or so (Hoekstra, Hyams and Becker, 1996; Chierchia, Guasti and Gualmini, 1999; Marinis, 1999; Pérez-Leroux and Roeper, 1999; Crisma and Tomasutti, 2000; Avrutin and Brun, 2002). Less documented is the fact that children's omission of articles varies across languages. Exceptions are the work by Lleó and Demuth (1999) (see also Lleó, 2001) and the work by Chierchia, Guasti and Gualmini (1999). Lleó and Demuth observed that there is a temporal displacement in the development of (proto)articles: children speaking Spanish produce proto-articles at an earlier age than German speaking children. Chierchia, Guasti and Gualmini (1999) observed that French and Italian speaking children cease to omit articles at an earlier point of linguistic development than English and Swedish speaking children. In this paper, we look at two new languages, Catalan and Dutch, beside Italian, and try to establish which factors influence children's omission of articles. The paper starts with a description of the system of articles and of the syntax/semantics of nouns in the three languages. Then, it presents the investigation of the speech of 9 children (3 for each language). It ends with a discussion of the results.

2. Background: the syntax, semantic and prosody of articles

As a first step let us briefly consider some basic facts concerning the cross-linguistic variation one finds with respect to the syntax of nouns in the three languages that we are going to consider, Catalan, Dutch, and Italian. Italian and Catalan have a full paradigm of articles: definite and indefinite articles varying in gender and number. In Italian, there are also some allophonic variants of the masculine articles (*lo, gli*); in Catalan and in Italian the definite singular feminine and masculine (*lo, el, la*) articles can be reduced to *l*. Dutch, instead, has three articles, two definite articles (*de* used with singular, plural, masculine and feminine nouns and *het* used with neuter nouns) and an indefinite one. Articles are used in front of common nouns in all three languages; in some varieties of Catalan (like the one under consideration), they must be also placed in front of proper nouns which are not vocative; in some varieties of Italian, proper nouns can be accompanied by an article. In addition, bare plural and mass nouns are also allowed, but their distribution varies to a significant extent. Bare plurals and bare mass nouns are grammatical in Dutch in all argumental positions. In Italian and Catalan bare plural and mass nouns are accepted only as sisters to a lexical head (the object of a transitive verb and the subject of an unaccusative verb) (Longobardi, 1998). This

cross-linguistic variation has been explained by assuming that the syntax of nouns is governed by the Nominal mapping parameter (Chierchia, 1998), a parameter that constrains the mapping between the syntactic category Noun and the corresponding semantic type. According to Chierchia, there are three types of languages. In one type, Nouns are mapped into Pred, in another they are mapped into Arg (and are thus names of kinds and have the typical behavior of mass nouns in Dutch or English) and, in the third type, Nouns can be either mapped into Pred or into Arg, a choice that is lexically determined. These mappings are instantiated respectively by the Romance languages, Chinese and the Germanic languages. This hypothesis allows one to explain a number of differences that these languages display. In Romance, Nouns are Pred and to turn them into arguments D must be projected and filled with an article. Thus, bare nominal arguments are generally disallowed in these languages. In Chinese, Nouns are born as Arg, they are names of kinds and have the syntactic properties of mass nouns. No article is needed and bare nominals can be used in all argument positions. Finally, Germanic languages are the union of the two types of languages just discussed. Nouns are either Pred or Arg. If they are Pred (count nouns) they behave like Romance nouns and need to be accompanied by an article; if they are Arg (mass nouns) they may occur as bare nominals. As for the way articles are prosodified, Catalan and Italian articles are proclitic to the following word. In Dutch, the articles *de* ('the') and *een* ('a') are weak forms and prosodify either as proclitic to the following word or as enclitic to the preceding word, often a verb. *Het* has a strong and weak form; in the former case, it prosodifies as a foot, while in the latter it cliticizes like the other Dutch articles (see Gussenhoven, 1985; Booij, 1995).

3. Data

Our investigation is based on the speech of 3 Catalan (Guillem, Laura and Pep), 3 Italian (Diana, Martina, Raffaello) and 3 Dutch speaking children (Abel, Peter, Thomas) (Catalan: from the corpus of M. Serra and R. Solé; Italian: Cipriani et al. 1989) all data being available through CHILDES (MacWhinney and Snow, 1985). Table 1 is a summary of the data used. It includes information about the files used, the age range during the period investigated and the MLU range. The MLU was calculated for all children on the first 100 utterances of each file used.

| | Files | Age range | MLU range |
|-----------|-------------------------------------|-----------|-----------|
| Guillem | 15,18,21,22,26,29,30,31, 32, 34 | 1;9-2;11 | 1.27-2.4 |
| Laura | 4, 6, 9, 11, 12 | 2;2-3;0 | 1.51-2.25 |
| Pep | 13,15,17,21,24,26,27,29, 31 | 1;5-3;0 | 1.2-2.95 |
| Diana | 1-9 (1, 3, 5, 8, 10) | 1;10-2;6 | 1.71-4.0 |
| Martina | 2-16 (2, 4, 6, 8, 11, 13, 16) | 1;11-2;4 | 1.25-2.49 |
| Raffaello | 1-14 (1, 2, 3, 5, 7, 9, 12, 13, 14) | 1;7-2;9 | 1.24-3.75 |
| Abel | 111,201,203,205,207, 210, 211, 301 | 1;11-3;0 | 1;29-3.6 |
| Peter | 111, 200, 202, 204, 205, 207, 208 | 1;11-2;08 | 1.8-2.8 |
| Tom | 110, 202, 205, 206, 209, 210, 301 | 1;10-3;01 | 1.17-3.13 |

Table 1: Data used in the study

We have eliminated unclear sentences, immediate repetitions of the same sentence, idiomatic expressions and routine sentences. Then, we have analyzed the relevant utterances by hand distinguishing between article use and omission in different contexts: in utterances including a verb and utterances without a verb (nouns in isolation, answers to questions or utterances not requiring a verb). This distinction is motivated by the following considerations. It is sometimes difficult to decide whether in utterances without verbs articles are needed or not. For example when one names an object or in lists and in answers to some questions, an article is not necessary. Thus, it is possible that in isolation omission of articles is overestimated. By contrast, in sentences, this problem does not arise. It is for this reason that we limited the analysis to verbal utterances. We counted as omissions of articles all nouns that in the target language would have required an article. This means that for Dutch, plural and mass nouns without an article were considered ungrammatical only in those contexts in which the use of an article was obligatory. Similarly, for Italian and Catalan, plural and mass nouns without an article were considered ungrammatical unless they occurred as sisters to a lexical head. Proper nouns without articles were counted as omissions for Catalan. Finally, in our counts we did not include some copular sentences where the distinction between argument or predicate was impossible to draw, as for sentences like “This tree is a mapple” or “John is a doctor”. This decision is motivated by the fact that predicates in these constructions can be used without an article. All the analyses that we have performed include only those files in which there were at least 5 relevant contexts for the use of articles.

4. Results

4.1 Age range

The age of children during the period investigated is roughly the same in the three languages (Catalan $M=29$, $SD=5.42$; Dutch $M=29$, $SD=4.46$; Italian $M=26$, $SD=4.63$). This ensures that any difference cannot be attributed to some children being younger than others.

4.2 Article omission in the three early languages

In order to compare the development of children’s omission of articles in the three languages, we had to match children’s performance based on some measure of linguistic development. We discarded age since children at the same age may be more or less advanced and there is enormous variability among children in the age range relevant here (Bates et al., 1995). As an independent measure of linguistic development we used the number of different words in the productive vocabulary. Therefore, we calculated for each child the number of different words produced in a given file and divided the data of article omission into 3 classes: the first class includes observations obtained when children employed between 0–100 different words, the second one when children used between 101–200 words and the third one when children used more than 201. Table 2 reports the means and SD of the

means of article omission at different stages of linguistic development in verbal utterances.

| | Stage 1 1-100 words | | Stage 2 101-200 words | | Stage 3 >200 words | |
|---------|---------------------|-----|-----------------------|------|--------------------|-----|
| | M | SD | M | SD | M | SD |
| Catalan | .45 | .48 | .06 | .004 | .01 | - |
| Italian | .52 | .24 | .17 | .19 | | |
| Dutch | .88 | .16 | .54 | .14 | .23 | .12 |

Table 2. Omission in different periods of linguistic development in utterances with verbs

A comparison of the means shows that at stage 1 there is no difference amongst the three languages. At stage 2, on the other hand, Catalan differs from Dutch ($p < .004$) and Italian differs from Dutch ($p = .05$). For stage 3, we have few data points, but we observe that omission is still present in Dutch, though not in Catalan. We can observe that by stage 2, omission has decreased to less than 25% in Romance, while it is still over 50% in Dutch. In summary, after a period in which omission is more or less over 50% in the three languages, there is a clear distinction between Romance speaking children and Dutch speaking children: the former omit less articles than the latter. In addition, at stage 2 Romance children are omitting half of the time of their Dutch peers, that is, they seem to have converged to the adult target system at a lower point of linguistic development, as measured by vocabulary size.

4.3 Age of first use of articles

We computed the mean age of first use of articles. Age of first use is defined as the age at which children use articles in a novel and non repetitive way. For each article we established the age at which this article was used for the first time. We distinguished between full forms of articles and phonetic approximations or reduced forms. We then computed the mean age of first use of full forms of articles and mean age of first use of full and reduced forms of articles. We should make precise that the first full form of articles was not found in the first file available, but in one of the subsequent files with the exception of *un* for Martina. The results are reported in table 3. We did not find any difference amongst the three languages when we considered only full forms of articles. By and large, all children start to use articles around age 2.

| | Age of first use (in months) | | Age of first use (plus protoarticles) | |
|---------|------------------------------|-----|---------------------------------------|-----|
| | M | SD | M | SD |
| Catalan | 26 | 4 | 22 | 2.8 |
| Italian | 24 | 2.8 | 22,6 | 4.5 |
| Dutch | 27 | 3 | 27,6 | 1.8 |

Table 3: Age of first use of articles (full forms and full forms plus protoarticles)

However, a difference was found when we included also protoarticles or phonetic approximations, which are essentially found in Italian and Catalan. Use of articles or phonetic approximations occurs earlier in Catalan and Italian than in Dutch. ($F(2,40)=8.21$, $p=.001$). Thus, children speaking a Romance language not only seem to converge earlier to the target system, but they start to use articles earlier. We turn next to consider which factors can explain article omission and which factors are at the heart of the difference found amongst the three languages.

5. Influences of input

5.1 Children omit because they find omissions in the input

We considered whether article omission is influenced by input factors: children may omit articles because omissions are found in their input. Thus, for each child, we examined the incidence of adult use of bare nouns at two points in the child's development, when the child omits many articles and when article omission has almost disappeared. Table 4 reports the results of this analysis.

| Child | Children's bare N | Adults' bare N | All Ns in adults |
|-----------|-------------------|----------------|------------------|
| Raffaello | 83% | 16% | 40% |
| | 0% | 06% | 27% |
| Martina | 83% | 27% | 29% |
| | 12% | 13% | 17% |
| Diana | 53% | 17% | 29% |
| | 8% | 19% | 37% |
| Thomas | 100% | 30% | 44% |
| | 7% | 21% | 38% |
| Abel | 68% | 33% | 43% |
| | 19% | 23% | 34% |
| Peter | 100% | 34% | 53% |
| | 28% | 18% | 30% |
| Guillem | 50% | 17% | 46% |
| | 0% | 19% | 31% |
| Pep | 100% | 23% | 38% |
| | 0% | 27% | 37% |
| Laura | 67% | 7% | 9% |
| | 2% | 26% | 28% |

Table 4: Omission of articles by children and their adult input

Let us describe how the table was obtained. For adults, we made two computations. First, we calculated the rate of bare nouns (be they grammatical or ungrammatical) over the total number of nouns. For Italian and Dutch, we only included common nouns, while for Catalan we also included proper nouns, since these require an article. We chose grammatical and ungrammatical bare nouns because they are part of the input (even beyond any conjecture about children's knowledge) and may be a model for article omission. In the second count, we computed the rate of bare nouns over all nouns including not only common nouns

but also proper nouns and nouns used in the vocative case (that do not require an article) for all the three languages. For adults, we considered nouns in utterances with verbs and in utterances without verbs. Table 4 shows that there are bare nouns in the caregivers' speech to children, but there is no decrease in the adults' use of bare nouns that parallels the decrease in the rate of bare nouns in child speech. This is not surprising since most of the cases of bare nouns in adult speech are grammatical. Thus, it seems unlikely that children omit and then cease to omit because they mimic the input. If this was the case, we should have found a change in the input (adults' speech) as we find a change in the output (children's speech).

5.2 Influences of the input on cross-linguistic differences in article omission

Although the incidence of bare nouns does not seem to be responsible for the children's article omission, it is possible that input factors are relevant for explaining the different development of article use in the three languages investigated. Remember that Romance learners seem to stop omitting articles at an earlier stage of linguistic development as measured by number of different words in the productive vocabulary. It is possible that Dutch children take more time to stop omitting articles because they are exposed to more adult cases of article omission. Therefore, we compared the overall rate of adults' bare nouns in the three languages (using the two files used in the previous analysis plus an additional one). We first computed the overall rate of bare nouns over the total number of nouns in three files chosen for the previous analysis. For Catalan, we included also proper nouns, for the reason explained above. We also calculated the overall rate of bare common and proper nouns and of nouns used in the vocative case over the total number of nouns.

We did not find any significant difference amongst the three languages. The rate of bare nouns in the caregivers' speech to children is by and large the same in the three languages, between 21% and 24% when we consider only common nouns (plus proper nouns for Catalan) and between 32% and 37% when we consider also proper nouns and nouns in the vocative case. Ungrammatical omissions of articles are rare in all three languages: 1% in Catalan, 2% in Italian and 4% in Dutch. These omissions are observed in cases in which the adult repeats an utterance without an article produced by the child; some rare cases are spontaneous uses of bare singular nouns used as proper nouns (*Cosa fa nonna?* 'What does grandmother do?'), especially for animals ('Look bear is eating') and for Dutch we found some cases of article omission with singular nouns that are not legitimate in the standard grammar, but acceptable in the spoken language in some special contexts. We return to these cases later.

The data discussed in the previous section and those presented in this section suggest that Dutch-speaking children do not continue to omit when Romance speaking children have ceased because they hear more bare nouns in the input. Thus, differences amongst the three languages do not depend on the rate of omission in the input to which subjects are exposed.

6. Prosodic constraints: The role of the prosodic properties of words

Lleó and Demuth (1999) claim that the timing of the appearance of articles is determined by prosodic constraints operating at the word level and by the way articles are prosodified in a particular language. They show that Spanish speaking children use articles earlier than their German speaking peers and suggest that this difference depends on the fact that Spanish learners have models for the prosodic integration of articles at the level of the vocabulary and, more specifically, in trisyllabic words with penultimate stress or with the prosodic structure WSW (weak-strong-weak). In trisyllabic WSW words the first weak syllable has the status of an unfooted syllable directly attached at the level of the prosodic word. Interestingly, in Spanish and more generally in Romance articles are prosodified with the subsequent word or are proclitics. They are weak unfooted syllables that, in combination with disyllabic words, give rise to the same prosodic structure of trisyllabic words with penultimate stress (WSW). This similarity in the prosodic structure of trisyllabic WSW words and of an article plus a SW disyllabic noun is at the basis of Lleó and Demuth's (1999) claim: the frequent availability of a model of prosodic structure with unfooted syllables at the lexical level (WSW trisyllabic words) boosts Spanish (and Romance) speaking children's production of articles. By contrast, such a prosodic model for the integration of articles is rarely available in German. Moreover, in German, articles are prosodified either as foot, if they are full forms, or as enclitic to the preceding word, often a verb, if they are reduced. These two facts, the rarity of a model and the manner of prosodifying articles, explains the German speaking children's reluctance to produce articles and is responsible for their later use with respect to Spanish.

We found that Romance speaking children start to use articles or phonetic approximations earlier than Dutch learners. We now evaluate whether this difference can be explained by appealing to prosodic constraints operating at the word level. The prosodic status of articles in Catalan and Italian is the same as the one of Spanish (articles are proclitic to the following word), while the prosodic status of Dutch articles bears some close resemblance to the one of German articles, as discussed in the background session. If prosodic constraints are responsible for the timing of appearance of articles, we expect that a model for the integration of articles is available in Catalan and Italian, while it is not (at least nor frequently) in Dutch.

Guasti and Gavarró (2003) showed that the prosodic properties of words in the early lexicon of Italian and Catalan are different. Catalan has quite a number of monosyllabic content words (27%), while Italian has almost none. By contrast, multisyllabic words (trisyllabic words or words with more than 3 syllables) are significantly more frequent in Italian than in Catalan (38% versus 19%). Thus, it is unlikely that trisyllabic WSW words are a model for Catalan speaking children for the integration of articles. However, Catalan might have another model for the integration of articles: disyllabic words with the stress pattern WS could be a model for the integration of articles with monosyllabic words.

Thus, we computed the rate of disyllabic WS words and of trisyllabic WSW words in the three languages under investigation. For this analysis, we chose a subset of the files we used for the analysis of article use. The proportions of

disyllabic WS words and of trisyllabic WSW words were entered into an analysis of variance with language (Catalan, Italian and Dutch) as a between factor and type of word (di- or trisyllabic) as a within factor. We found a main effect of language ($F(2,6)=34.02, p=.0005$), of type of word ($F(1,6)=93.52, p<.0001$) and an interaction between language and type of word ($F(2,6)=124.27, p<.0001$). The interaction is due to the fact that in Italian and Dutch WS disyllabic words are very infrequent (less than 1%), while they are frequent in Catalan (41%). Trisyllabic WSW words are frequent in Italian (81%), but not in either Catalan or Dutch (23% and 18% respectively). A post hoc Scheffé test on the factor language reveals that Catalan is different from Dutch ($p=.004$) and Italian is different from Dutch ($p<.001$). From this we infer that there is indeed a model for the prosodic integration of articles in the vocabulary of Catalan and Italian. It is possible that learners of Catalan and Italian start to use articles earlier than learners of Dutch because they are more familiar with the possibility of having unfooted syllables at the lexical level.

Both disyllabic and trisyllabic words with the relevant stress pattern are models for the integration of articles in Catalan and possibly in Dutch. Therefore, we might expect that the timing of appearance of articles is inversely proportional to the frequency of the model: the more a model is frequent the sooner the child will start to use articles. Thus, we performed a second analysis in which we put together WSW and WS words for Catalan and Dutch but only at stage 1. We did not collapse WS and WSW words in Italian; here we only used WSW words as a model: in Italian there are no monosyllabic content words so WS words cannot be a model for the integration of articles in Italian. We entered the data into an analysis of variance with language as a factor. We found a main effect of language ($F(2,6)=117.21, p<.0001$). A post hoc Scheffé test showed that there is a significant difference between Italian and Catalan ($p=.0002$), Italian and Dutch ($p<.0001$) and Catalan and Dutch ($P=.003$). The model for the integration of articles is very frequent in Italian (86%), less so in Catalan (36%) and even less in Dutch (7%). If frequency of a model was the relevant factor for the bootstrapping of the use of articles, we would expect articles to be used earlier in Italian than in Catalan and earlier in Catalan than in Dutch. But we did not find any difference between Italian and Catalan. This suggests that the frequency of the model is an important factor, as Dutch speaking children start to use articles later than their Romance speaking peers, but probably it acts more as a threshold: a certain amount of a pattern (let's say at least 36%) is enough to bootstrap the use of articles. A higher frequency of exposure to the model does not grant better results.

7. Discussion

Although all children omit articles in their first multiword combinations, the developmental course of article use is subject to cross-linguistic variation. First, articles or their phonetic approximations start to be used earlier in child Romance than in child Dutch. Second, while at stage 1, the rate of article omission is by and large the same in the three languages investigated, at stage 2 there is a clear difference: Catalan and Italian speaking children use articles much more than their Dutch peers. Thus, we could suggest that Romance learners reach a high level of article use (around 75%) at a lower point of development when compared to Dutch

speaking children (a finding that replicates what was found in Chierchia, Guasti and Gualmini (1999) with a different set of languages). These findings raise two questions: (a) why do children omit articles? and (b) why does article omission display different patterns in early languages? We established that children's omission of articles is not input driven. Adults use bare nouns in their child directed speech, but this use is not subject to changes during the course of development. Children hear more or less the same rate of bare nouns over the developmental course examined here. Still, their use of articles increases and this is not guided by the input. Second, we also established that the rate of bare nouns in the caregivers' speech is not different in the three languages investigated. This fact discards the possibility that differences among the three languages are determined by the rate of bare arguments in the adult speech. We then examined whether prosodic constraints can play a role in the use of articles. According to Lleó and Demuth (1999) prosodic constraints operating at the lexical level in combination with the way articles are prosodified in a particular language predict the appearance of articles. More specifically, they claimed that it is the availability at the lexical level of a model for the prosodic integration of articles that boosts the children's use of articles. We established that both in Italian and Catalan there are models for the integration of articles in trisyllabic words with penultimate stress (WSW words) and in disyllabic iambic words (WS words), respectively. By contrast, these models are less frequent in Dutch. These results lead us to conclude that prosody may indeed boost the children's use of articles in Romance and thus explain some of the cross-linguistic differences in article use. Furthermore, it could be argued that it is because Romance learners start to use articles or phonetic approximations earlier than they also converge at a lower point of linguistic development: they have had more experience with articles. This conjecture leaves some of our findings unexplained, however. First, it is not the case that the more a model is frequent, the earlier children start to use articles. If this was the case, we would expect Italian children to start using articles earlier than Catalan, but this does not happen. Second, the frequency of exposure to a model does not predict the course of development of article use. We looked at the developmental course of words with the stress pattern WSW in Dutch in the three children examined and we did not find a clear increase of that pattern parallel to the decrease of article omission. For example, when Thomas omits articles 98% of the time, he uses 33% of WSW trisyllabic words over the total of trisyllabic words and when article omission has decreased to 11% WSW trisyllabic words represent 21% of all the trisyllabic words. Similar observations can be extended to disyllabic words. Thus, it is not the case that we observe an increase of the model at the lexical level parallel to the decrease of article omission and this is so in all the three early languages. Finally, although Romance learners start to use articles earlier than Dutch learners, they still omit quite a lot of articles in the first stage of linguistic development.

7.1 The Nominal mapping parameter

Our results are compatible with an account of the acquisition of articles in terms of the Nominal Mapping Parameter (NMP) (Chierchia, 1998) discussed in session 1. In stage 1, we did not find any difference in article omission amongst the three early

languages. Children's grammar at this point of development can be described as the result of having chosen the Germanic setting of the NMP, that is, children take nouns to be either Pred or Arg. By assuming this setting, children have to decide noun by noun whether it is to be mapped into Arg (it is mass) or Pred (it is count). Omission of articles is the result of incorrectly mapping a noun into the semantic type Arg. Thus, article omission in the first stage is the result of the Germanic setting plus the fact that this setting requires the child to decide whether a noun is an Arg or a Pred and she may be a bit confused in operating the classification. When Romance learners discover that the German setting is not valid for their language of exposure, they switch the value of NMP to the category Pred. This seems to happen at stage 2, when Italian and Catalan diverge from Dutch. This change may be triggered by different aspects of the article system: the fact that the article paradigm is complete in Romance (there are definite, indefinite, singular and plural articles); the use of expletive articles, that is, articles with proper nouns (see Marinis, 1999), which in Catalan are obligatory and in Italian are at least optional. At stage 2, Romance learners have converged to the adult grammar and omission decreases to under 25%. Dutch learners still omit articles at a higher frequency because the German setting of the NMP requires a classification noun by noun and this decision may take time.¹ However, this might not be the only reason. In fact, also the Catalan and Italian learners have to distinguish between count and mass nouns, since bare plural and mass nouns are allowed in the sister position of a lexical head in Romance, as we showed in the background section. Still, this lexical classification does not lead Catalan and Italian learners to omit articles at stage 2 at a high rate; in fact at stage 2 they can be considered to have converged to the adult system. Therefore, it is possible that some other factor is contributing to the Dutch children's omission of articles.

7.2 Special registers or abbreviated languages

Although not in the standard variety, Dutch allows singular nouns to be used without an article in special registers. At the spoken level, we find examples with bare singular nouns, both in subject and object position, with several also present in our corpus (although their frequency was not high, 3–4%). Some examples from the corpus examined are reported below:

- (5) Dat is een meisje van twee huizen verderop. Meisje van een jaar of zes / zeven. (Peter 20419)
This is a girl who lives nearby. Girl aged six/seven year old.

Sentences like the ones above are fully acceptable despite the fact that they seem to violate some apparently syntactic constraints. However, they are acceptable only when the speaker and the listener share some piece of knowledge, either

¹ The problem is aggravated by the fact that mass and plural nouns can be used without an article, but this is subject to some constraints. In *The furniture is from Italy* the article cannot be omitted. So, children also have to figure out the contexts in which mass and plural nouns can be used without an article.

because they are talking about someone familiar to both or because they are discussing something that has happened in the context and that was witnessed by both. Under the same pragmatic conditions, examples like those above are not acceptable in Catalan and Italian. Let us return to why Dutch learners are still omitting at stage 2 at a significantly higher rate than Catalan and Italian learners. One possibility is that Dutch learners omit articles also because they use the grammar of special registers without yet knowing the pragmatic conditions that legitimate its use, as suggested by Baauw et al. (2002) and Avrutin (in press). Dutch special registers extend the standard grammar of Dutch in that they allow bare singular nouns in subject and object position. As we alluded to above, in the spoken language, omission of articles is not allowed in Italian and Catalan. In these languages we do find bare nouns in headlines and advertisements (a possibility also found in Dutch):

- (6) Empresa familiar catalana cerca model de creixement
business family Catalan looks for model of development
'Catalan family business looks for model of development.'

As in Dutch special registers, the noun without an article must be familiar or refers to some event or individual that is fairly specific, although not definite. One common aspect of the grammar of special registers in Catalan, Dutch and Italian is that bare nouns are allowed under similar pragmatic conditions: the noun must be familiar. However, the syntactic properties of these bare nouns are different. In Catalan and Italian special registers, bare nouns are found in utterance initial position most of the time (88% in Italian and 94% in Catalan, based on a corpus of 100 headlines we have collected); in Dutch, by contrast, bare nouns can be in initial or object position (in initial position of headlines we have found 61% bare nouns). Moreover, special registers are restricted to headlines and advertisements in Catalan and Italian, while in Dutch they are part of the normal conversation and part of the input that children receive and thus may confuse Dutch learners.

In summary, we argue that the prosodic properties of words can help Catalan and Italian children in bootstrapping the use of articles. But once children have discovered that articles are to be used, they still have to figure out the syntactic properties of nouns in their language. Regardless of the language they are exposed to, children omit articles for a while and this can be accounted for by assuming that they have adopted the German setting of the Nominal Mapping Parameter. Therefore, Nouns are taken to be Arg or Pred. If they are mapped into Arg, they can be used without an article. Romance children figure out that Nouns must be mapped to Pred and stop omitting articles. Dutch children continue to omit articles at a point of linguistic development when Romance children have considerably decreased their article omission. We attribute this to two factors. The German setting of the Nominal Mapping Parameter has a lexical component: children have to decide noun by noun if it is to be mapped into Pred (count noun) or into Arg (mass noun) and this decision takes some time. In addition, they have to figure out the pragmatic conditions operating in special registers and allowing omission of articles with bare singular nouns.

Acknowledgments

The work of Joke de Lange on this publication was supported by the project Comparative Psycholinguistics, which is funded by the Dutch Organisation for Scientific Research (NWO). The work of Anna Gavarró received the financial support of the MCyT through project BFF2000-0403-C02-02.

References

- Avrutin, S. (in press) 'Beyond narrow syntax', in: L. Jenkins (ed.) *Genetics of Language*. London: Routledge
- Baauw, S., E. de Roo and S. Avrutin (2002) 'Determiner omission in language acquisition and language impairment: Syntactic and discourse factors', in: A.H.-J. Do, L. Dominguez and A. Johansen (eds.) *Proceedings of the 25th BUCLD*. Somerville MA: Cascadilla Press, 24-35.
- Bates, E., P. S. Dale and D. Thal (1995) 'Individual differences and implication for theories of language development', in: P. Fletcher and B. MacWhinney (eds.) *Handbook of Child Language*. Oxford: Blackwell.
- Booij, G.E., (1995) *The Phonology of Dutch*, Oxford: Clarendon Press.
- Chierchia, G. (1998) 'Reference to kinds across languages'. *Natural Language Semantics* 6, 339-405.
- Chierchia, G., M.T. Guasti and A. Gualmini (1999) 'Nouns and articles in child grammar and the syntax/semantics map'. Paper presented at GALA, Potsdam, September.
- Cipriani, P., A.M. Ciliosi, P. Bottari, and L. Pfanner (1993) *L'acquisizione della morfosintassi: Fasi e processi*. Padova: Unipress.
- Crisma, P. and E. Tomasutti (2000) 'Phonological effects on article omission in the acquisition of Italian', in S. C. Howell, S.A. Fish and T. Keith-Lucas (eds.) *BUCLD 24 Proceedings*. Somerville MA: Cascadilla Press, 220-231.
- Guasti, M. T. and A. Gavarró (2003) 'Catalan as a testing hypothesis concerning article omission'. *Proceedings of the 27th BUCLD*. Somerville MA: Cascadilla Press, 288-298.
- Gussenhoven, C. (1985), 'Over de fonologie van Nederlandse clitica', *Spektator* 15, afl.3, 185-200.
- Hoekstra, T., N. Hyams and M. Becker (1996). 'The underspecification of Number and the licensing of root infinitives', in: E. Hughes et al. A. Stringfellow, D. Cahana-Amitay, E. Hughes and A. Zukowski (eds.) *BUCLD 20 Proceedings*. Somerville MA: Cascadilla Press, 293-306.
- Lleó, C. and K. Demuth (1999) 'Prosodic constraints on the emergence of grammatical morphemes: crosslinguistic evidence from Germanic and Romance languages', in: A. Greenhill, H. Littlefield and C. Tano (eds.) *Proceedings of the 23rd BUCLD*. Somerville MA: Cascadilla Press, 407-418.
- MacWhinney, B. and C. Snow (1985) 'The Child Language Data Exchange System'. *Journal of Child Language* 12, 271-296.
- Marinis, T. (1999) 'Minimal inquiries and acquisition of the definite article in Modern Greek'. *Proceedings of the 34th Colloquium of Linguistics*. Frankfurt: Peter Lang Verlag, 73-82.
- Pérez-Leroux, A.-T. and T. Roeper (1999) 'Scope and the structure of bare nominals: Evidence from child language'. *Linguistics* 37, 927-960.