Article Omission: Regularization and Generalization

Andrea Ceolin (ceolin@sas.upenn.edu) – University of Pennsylvania

Introduction

Children speaking Romance languages can produce articles earlier than children speaking Germanic languages (Lleó and Demuth 1999, Kupisch 2007, Guasti et al. 2008). Previous works deny any role of the input in predicting cross-linguistic differences. We challenge this claim by showing that children learning English are less exposed to articles than Italian and Spanish ones, using data from CHILDES.

Prosodic account

Lleó and Demuth (1999) note that at approximately 21 months German children omit articles 90% of the time, while at the same age Spanish children omit articles only 40% of the time.
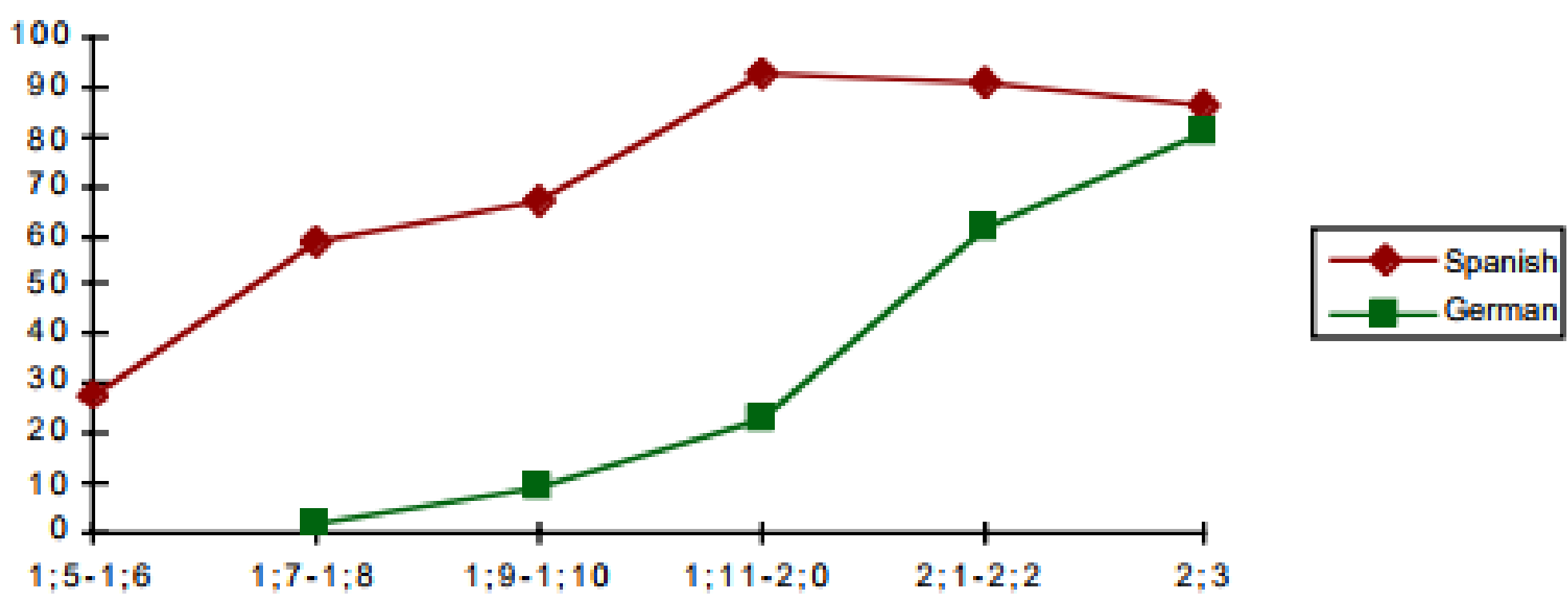

They also show that Spanish children at the same age can produce trisyllabic words, while German children cannot. They explain this phenomenon by showing that multisyllabic words in the early target vocabularies are more frequent in Spanish than in German.
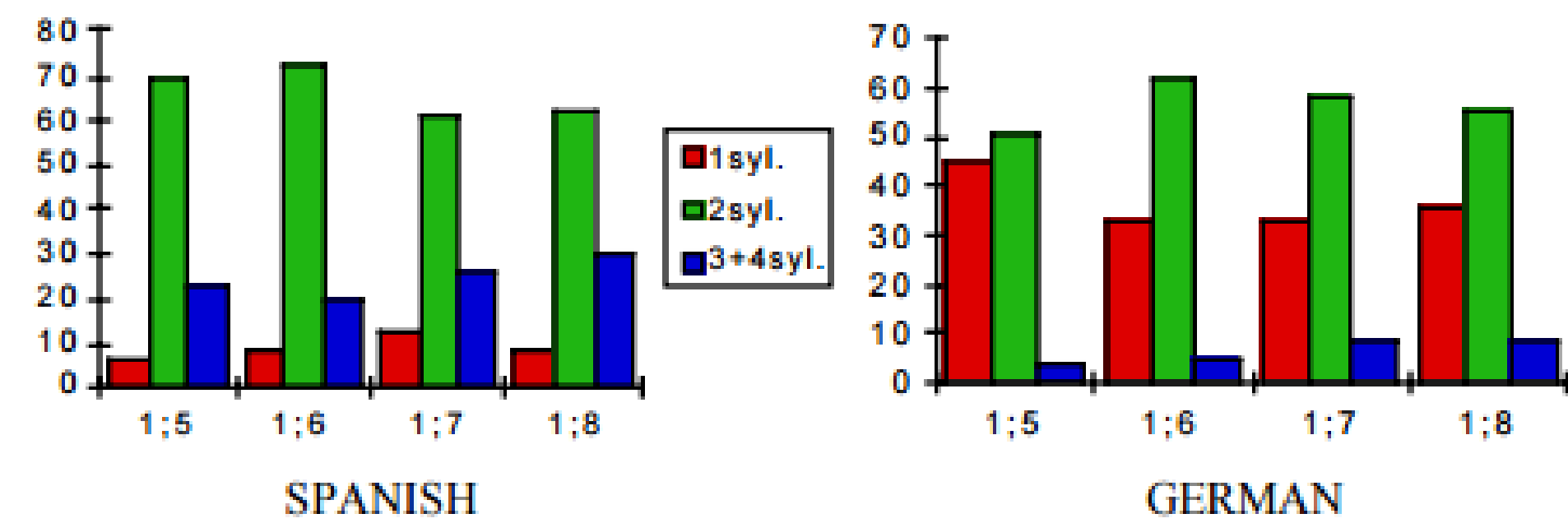
Additionally, experimental evidence shows that learning how to produce multisyllabic words is not sufficient to integrate articles consistently. Consider the following experiment (1a-b) in Gerken (1996) on 26-month-old children. If articles cannot be included in a S-w foot, they are often dropped. This is also true for 30-month-old Italian children (2a-b, from Crisma and Tomasutti 2000).

Figure 1 – Frequency of articles (blue) and possessives (red) in Italian, Spanish and English noun phrases. The top-50 most frequent nouns are considered. Data from CHILDES: Calambrone (Italian), Ornat and OreaPine (Spanish) and Brown (English)
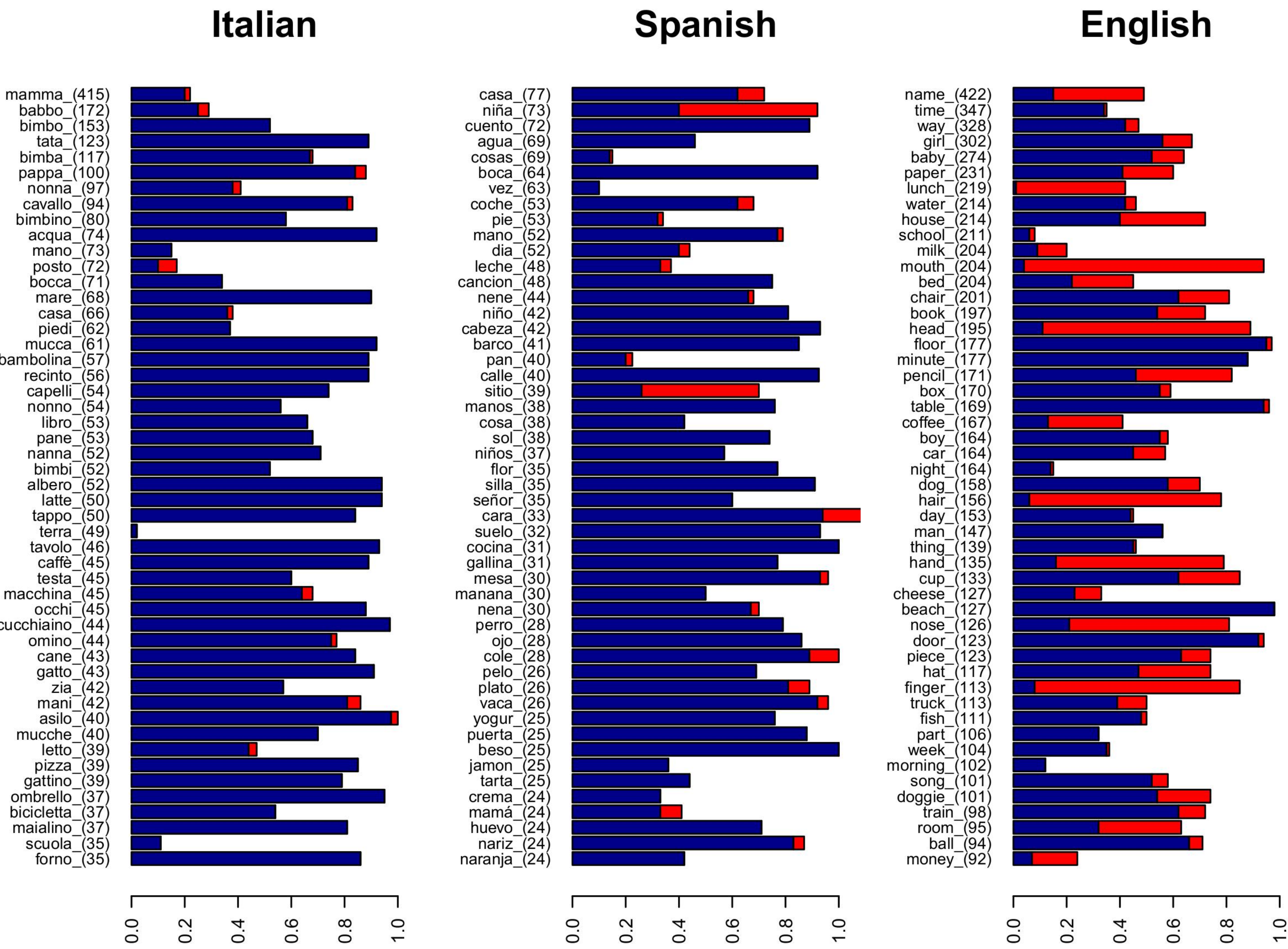


```
      *    S    w   S-(w)
1a. He [KICKS the] [PIG]           (omission rate: 11%)

      *    S    w    *   S-(w)
1b. He [CATCHes] the [PIG]         (omission rate: 46%)


     S         w   S   w
2a. [FA      la] [TORta]           (omission rate: 8%)
    Make.3SG the  pie

     S    w    *   S   w
2b. [PRENde]  la [TORta]           (omission rate: 83%)
    Take.3SG    the pie
```

Some questions remain open. First, Spanish children produce articles even before they can utter trisyllabic words, while German ones do not. Second, the theory predicts a low omission rate for languages like French, but this is not attested (Kupisch 2007). On the other hand, languages like Catalan are falsely predicted to have a high rate of omission (Guasti et al. 2008). The prosodic account does not explain the uniform behavior of children speaking Romance.

Nominal Syntax

We found that children are not exposed to the same amount of definite articles cross-linguistically. In particular, the syntax of possessive determiners plays a crucial role:

▪ **Saxon genitives**: Germanic languages can license possessors through a -s clitic/suffix. In all these instances, Romance languages uniformly exhibit a definite article.

▪ **Postnominal possessives**: Det + N + Poss cannot be used for definite readings in Germanic languages, while it is available in Romance (Bernstein 2005).

▪ **Inalienable possession**: Romance languages use definite articles to express inalienable possession, contrary to languages like English and Dutch.

Figure 1 shows that English noun phrases contain less articles (blue bar) than Italian and Spanish ones, because of their competition with possessives (red bar). Since possessives can be used with mass/count nouns and in argument/non-argument position, they are not useful to learn the conditions under which articles are obligatory.

Regularization

In the presence of inconsistent input, children acquire a rule if it appears 40% of the time (Newport 1999, Schuler 2017). This means that nouns which appear with a definite article at least 40% of the time qualify for learning the rule that articles are obligatory. On the other hand, nouns in which definite articles are rare can be treated as exceptions.
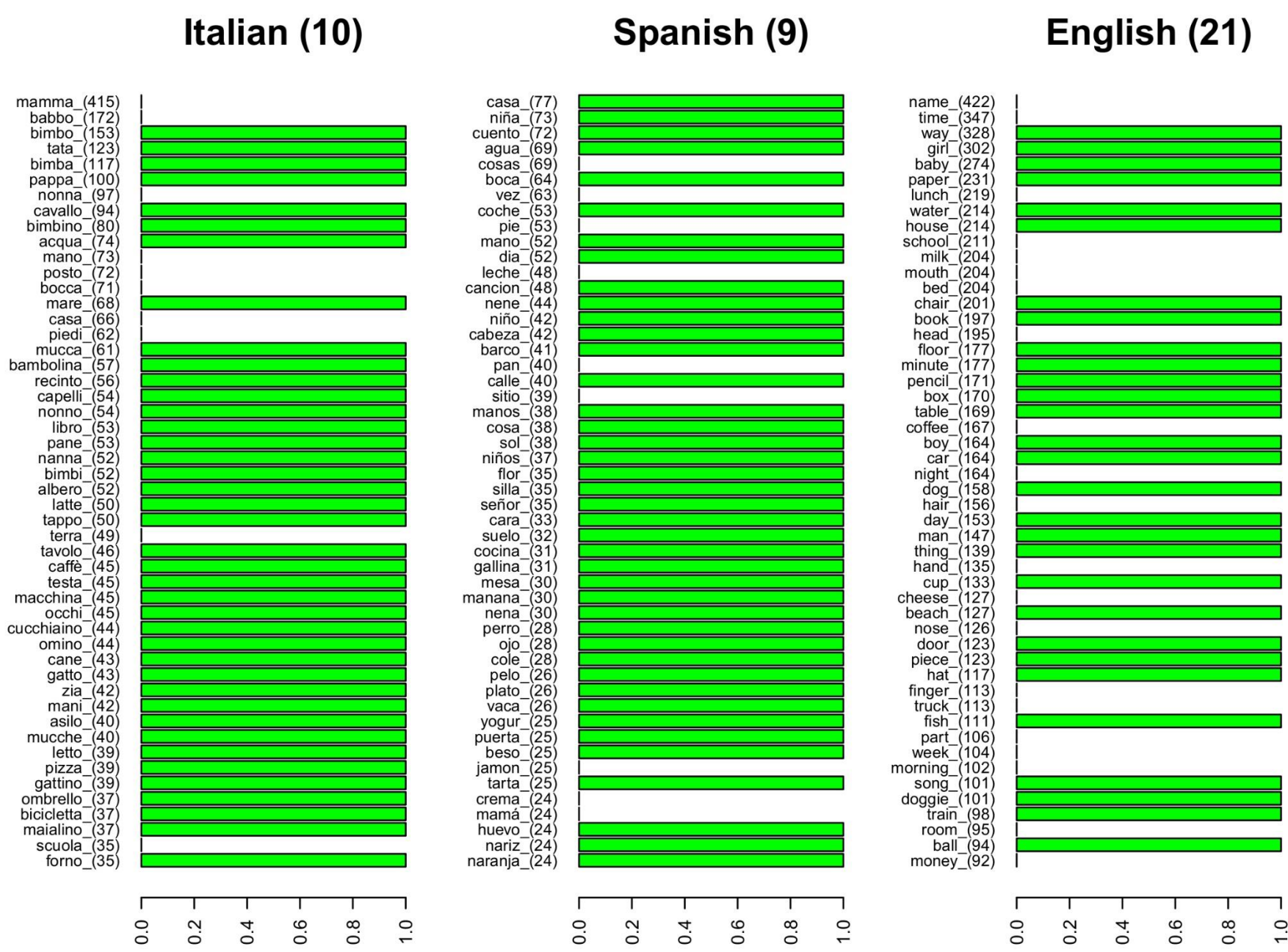
Generalization

According to the Tolerance Principle (Yang 2016), a rule can generalize to N items if the amount of exceptions does not exceed $\theta = N/\ln(N)$. Over a vocabulary of 50 words, approximately 13 exceptions can be tolerated, a number which is exceeded by English. Article omission in English can be linked to a failure in generalizing its obligatory use, which might lead children to temporarily consider articles as optional in absence of robust evidence (Valian 1991).

References and Acknowledgements

Bernstein, J. B. 2005. On the morpho-syntax of possessive constructions. Recherches linguistiques de Vincenns (34).

Crisma, P. & Tomasutti, E. 2000. Phonological effects on article omission in the acquisition of Italian. Proceedings BUCLD 24.

Gerken, L. 1996. Prosodic structure in young children's language production. Language.

Guasti, M. T., Gavarro, A., De Lange J. & Caprin C. 2008. Article omission across child languages. Language Acquisition 15 (2).

Kupisch, T. 2007. Testing the effects of frequency on the rate of learning: Determiner use in early French, German and Italian. Proceedings of the workshop on input frequencies in acquisition.

Lleó, C. & Demuth, K. 1999. Prosodic constraint on the emergence of grammatical morphemes: Crosslinguistic evidence from Germanic and Romance languages. Proceedings BUCLD 23.

Newport, E. L. 1999. Reduced input in the acquisition of signed languages: Contributions to the study of creolization.

Schuler, K. D. 2017. The Acquisition of Productive Rules in Child and Adult Language Learners. Doctoral dissertation.

Valian, V. 1991. Syntactic subjects in the speech of young children. Developmental Psychology 22(4).

Yang, C. 2016. The price of linguistic productivity: How children learn to break the rules of language. MIT Press.