

CONSONANT INVENTORIES IN THE SPONTANEOUS SPEECH OF YOUNG
CHILDREN: A NEW PROCEDURE

Lieve Van Severen *

Renate van den Berg *

Inge Molemans *

Steven Gillis *

* University of Antwerp, Department of Linguistics, CLiPS Research Centre

Address for correspondence:

Steven Gillis

University of Antwerp

Department of Linguistics, CLiPS Research Centre

Prinsstraat 13

B-2000 Antwerpen

Belgium

Email: steven.gillis@ua.ac.be

CONSONANT INVENTORIES IN THE SPONTANEOUS SPEECH OF YOUNG
CHILDREN: A NEW PROCEDURE

Abstract

Consonant inventories are commonly drawn to assess the phonological acquisition of toddlers. However, the spontaneous speech data that are analysed often vary substantially in size and composition. Consequently comparisons between children and across studies are fundamentally hampered. The current study aims to examine the effect of sample size on the resulting consonant inventories. A spontaneous speech corpus of 30 Dutch-speaking two-year-olds was used. The results indicate that in order to construct and compare inventories reliably, they should be drawn from speech samples that are equally large. A new consonant inventory procedure is introduced. The implementation of this procedure demonstrates considerably less variation in inventory size across children and word positions than reported previously. This finding has important implications for clinical studies that constructed and compared inventories of typically and atypically developing children without normalizing the sample size.

Key words

phonological acquisition, child language, methodology

Introduction

A consonant inventory is defined as the set of consonantal types that occur in a child's production of meaningful words at a particular moment in development. Traditionally, two methods have been applied: 1) independent analyses, in which all consonantal types with phonemic status in the target language that occur in the actually produced word forms are listed without reference to the adult target form (cf. Stoel-Gammon, 1985; 1987; Dinnsen, Chin, Elbert & Powell, 1990; Stoel-Gammon, 2002; Stokes & To, 2002), and 2) relational analyses, in which the accuracy of the consonant productions in relation to the target forms is considered in the sense that only correctly produced consonants are used in the analysis. (cf. Templin, 1957; Sander, 1972; Beers, 1995; Stokes, Klee, Carson & Carson, 2005). Thus, for instance, if a child produces the English word "deep" as [dit], only the segment [d], and not the final [t], will be considered in a relational analysis because the [t] is a substitute for [p], while in an independent analysis initial [d] as well as final [t] are considered. The analyses reported in this paper are restricted to independent analyses.

The aim of consonant inventory studies is to chart the normal developmental path of the acquisition of segments and to assess the variation present in a typically developing population. Because the size and the segmental composition of 'fully developed' consonant inventories differ between languages, group data have already been provided for a variety of languages, such as English (Stoel-Gammon, 1985; 1987; Dyson, 1988; Robb & Bleile, 1994; Watson & Scukanec, 1997; Stoel-Gammon, 2002), Finnish (Kunnari, 2003; Savinainen-Makkonen, Kunnari & Paavola, 2008), Italian (Zmarich & Bonifacio, 2005), Arabic (Amayreh & Dyson, 2000), Quiché (Pye, Ingram & List, 1987) and Cantonese (Stokes & To, 2002). These data allow an investigation of whether and to what extent developmental paths are idiosyncratic or, alternatively, show a universal pattern (Pye, et al., 1987; Fikkert, 1994; Amayreh & Dyson, 2000; Zamuner, 2003; Catano, Barlow & Moyna, 2009). Moreover, the

developmental data of clinical populations, such as children with specific language impairment, can be charted against these typical profiles (Paul & Jennings, 1992; Thal, Oroz & McCaw, 1995; Rescorla & Ratner, 1996; Carson, Klee, Carson & Hime, 2003).

Constructing and comparing consonant inventories is not without difficulties. The size and the segmental content of consonant inventories can be influenced by the choices that are made in the selection of the speech material. These choices concern the nature of the data (e.g. whether or not meaningless, unintelligible, imitated and/or repeated utterances are selected) as well as the amount of speech (e.g. how many minutes, utterances, word types and/or variants of the same word type are selected?).

Obviously, the size of a consonant inventory depends on the size of the speech sample from which it is drawn: the larger the sample, the more different segments can potentially be found. Hence inventory studies should use equally large data sets for all individuals and groups that are compared. Investigators have often implemented this requirement by selecting recordings of equal duration, e.g. a recording of 10 or 20 minutes (Rescorla & Ratner, 1996; Carson et al., 2003). However, ‘a particular amount of time’ may not be appropriate, because the amount of speech within a particular time period can vary substantially, depending on the personality, the general arousal of the child, the period of the day, the type of interaction, etc. (Bornstein, Painter & Park, 2002). If the size of the speech samples can differ between children and across time, the resulting consonant inventories may vary between children and across time as well. For instance, Morris (2009) collected two recordings of her subjects of exactly the same duration within the same week. The recordings consisted of a substantially different number of words, and the resulting consonant inventories substantially varied in size as well. It is unclear whether and how the size and the composition of the speech sample are linked to the resulting consonant inventory.

In the literature, the influence of the nature and the amount of speech data on commonly used phonological measures, such as consonant inventories, has hardly been investigated (but see Tomasello & Stahl, 2004; Edwards & Beckman, 2008; Rowland, Fletcher & Freudenthal, 2008; Morris, 2009). To contribute to this area, the current investigation aims to:

- (A) determine the exact relation between the amount of speech data and the size of the consonant inventory.
- (B) evaluate various units of speech selection (amount of time versus linguistic units) for the construction of consonant inventories.
- (C) determine the required sample size for constructing consonant inventories reliably.
- (D) propose a new consonant inventory procedure that fulfils the methodological requirements concerning the amount of speech and the unit of analysis.

Throughout this study one large data set is used. Spontaneous speech of 30 typically developing children acquiring Dutch (aged between 1;10 and 2;0). Because Dutch has 21 consonants in word-initial (WI) position and 14 consonants in word-final (WF) position (Booij, 1995), separate inventories will be established for WI and WF consonants: WICIs (word-initial consonant inventories) and WFCIs (word-final consonant inventories).

Method

Participants and data collection

Thirty typically developing children acquiring Dutch (14 girls and 16 boys) between age 0;6 and 2;0 participated in the study. They were recruited on the basis of the following selection criteria: living in the Dutch-speaking part of Belgium, no health or developmental problems, normal hearing, no twins, monolingual and no repeated scores less than percentile 1 on the N-

CDI, the Dutch version of the MacArthur Communicative Development Inventories (Zink & Lejaegere, 2002) at age 1;0, 1;6 and 2;0. The parents of the children spoke standard Dutch and had a mid-to-high socio-economical status. The investigation was institutionally approved and the parent(s) signed a written consent.

Video- and audio recordings of naturalistic interactions between the child-parent dyads were made at the children's homes on a monthly basis. For each observation, 50 minutes to 3 hours of speech were recorded with a JVC digital camera (type GZ-MG77E) and a built-in, multidirectional microphone. Following the recording session, the investigator edited a selection of 20 minutes of speech interaction between the vocally active child and the parent(s) or the investigator. In this particular study, for each child data of three successive sessions at 1;10, 1;11 and 2;0 were merged and were denoted by the term entire speech session.

Data transcription

Three transcribers made orthographic and phonemic transcriptions of the lexical productions in CLAN, according to the CHAT conventions (MacWhinney, 2000). The procedure proposed by Vihman & McCune (1994) was applied to identify meaningful speech. All 22 Dutch consonants with phonemic status, including consonants that occur in loan words, were transcribed (Booij, 1995).

The reliability of the transcriptions was computed on 15% of the entire speech corpus. A dynamic alignment algorithm based on ADAPT (Elffers, Van Bael & Strik, 2005) was implemented to align the phonemic characters of the transcription pairs. The transcribers' agreement on the identity of the consonants is acceptable: the percentage of intrarater agreement is 84.17% (Cohen's Kappa = 0.83), and the percentage of interrater agreement is 70.43% (Cohen's Kappa = 0.68).

Data characteristics

Table 1 demonstrates the extent of individual variation in volubility. The median and the range computed over all children are shown for four linguistic units: the number of word tokens and types, and the number of WI and WF consonant tokens. Although in each case speech selections of equal duration were extracted from the original recordings, the data in table 1 reveal substantial individual differences. The difference between the size of the entire speech session of the least and the most voluble child amounts to 1438 word tokens, 230 word types, 1105 WI and 743 WF consonant tokens.

[INSERT TABLE 1 AROUND HERE]

Bootstrapping procedure

Table 1 shows substantial variation in the number of words and consonants. Nonetheless, samples of equal size are required to reliably construct and compare the consonant inventories drawn from them. Therefore, for each child a ‘sub’sample of a fixed size N was selected from the entire speech session. The sample was randomly selected to exclude a sampling bias. For instance, if a sample of continuous, spontaneous speech is selected in which the child is only talking about the ball he/she is playing with, the consonant inventory may be substantially smaller than when the speech is randomly selected from various situations (e.g. playing with the ball, looking in a book, helping daddy in the kitchen, etc.). Furthermore, the entire session includes more speech than the speech selected in the sample. If another random sample would have been selected, perhaps a slightly different consonant inventory (size) could have been attested. To normalize for this effect of sampling, not one

single sample was taken, but 1000 unique samples were randomly drawn from the session. Then the average consonant inventory size or the mean incidence of a single consonant was computed over all 1000 samples. As a result, the variation due to sampling was taken into account. The technique of repeatedly drawing random speech samples is called a ‘bootstrapping’ or ‘Monte Carlo’ procedure (Efron, 1979; Baayen, 2008), and was also applied by other researchers investigating lexical and grammatical aspects of child language development (McKee, Malvern & Richards, 2000; Taelman, Durieux & Gillis, 2004; Xanthos & Gillis, 2010). The Monte Carlo procedure is implemented in all analyses reported in this paper using a Perl-script.

Data analysis

(A) Determine the exact relation between the amount of speech data and the size of the consonant inventory

In order to investigate the relation between the amount of speech data and the size of the consonant inventory, inventories were drawn from speech samples of varying size. When selecting the speech material and establishing the consonant inventories, the instructions of a commonly applied procedure were followed (i.a. Stoel-Gammon, 1985; 1987; Dyson, 1988; Kunnari, 2003; Morris, 2009). More specifically, the procedure used in the most recent study (Morris, 2009) was implemented. According to Morris, the following speech selection criterion and consonant inclusion criterion can be considered as standard practice:

- Speech selection criterion: Only intelligible words and the first two variants of a word type are selected from the entire speech session.
- Consonant inclusion criterion: All consonants that occur in at least two different word types are added to the inventory.

The flowchart in figure 1 demonstrates the procedure that was followed for each child. The procedure consists of five (conditionally linked) steps and was implemented in a Perl-script.

Broadly speaking the following steps were taken:

- Step 1: Only those word productions fulfilling the word selection criterion are selected from the entire speech session. The resulting selection consists of K word tokens that are used in the following steps.
- Step 2: From this material, a speech sample s_i of N word tokens is randomly drawn. Initially, N is (arbitrarily) set to 20 word tokens.
- Step 3: The consonant inventory, i.e. the number of different consonantal types in the speech sample, is determined according to the consonant inclusion criterion.
- Step 4: 1000 speech samples with sample size N are drawn. Each sample is denoted by s_i ($0 < i < 1000$).
- Step 5: The average consonant inventory size is computed for the 1,000 samples.

Sample size N is increased systematically with an arbitrarily chosen step-size of 15 words.

Then, Steps 2 to 5 are repeated until no sufficient data are available to select at least 1000 unique combinations of N word tokens from the speech selection. The number of potential unique combinations is computed by means of the binomial formula (1),

$$(1) \quad C(N,K) = \frac{N!}{(K!(N-K)!)}$$

where N stands for the number of word tokens in the sample and K denotes the number of word tokens in the entire speech material.

[INSERT FIGURE 1 AROUND HERE]

(B) Evaluate various units of speech selection

Does the unit of speech selection affect the size of the consonant inventory? Four procedures were compared that only differed with respect to the unit of selection: a particular amount of time or a particular linguistic unit. If time was the unit of speech selection, then recordings of equal duration were analysed. If a linguistic unit was the unit of selection, then an equal number of utterances, or words, or segments was used for compiling the consonant inventory. Because a consonant inventory procedure that uses the consonant as unit has not been proposed yet, a new procedure was introduced. To allow for comparisons, this procedure was applied for other units as well. For all units of speech selection, the consonant inventory procedures used the same word selection and consonant inclusion criteria:

- Speech selection criterion: all intelligible word tokens that are produced by the child are selected.
- Consonant inclusion criterion: a stringent inclusion criterion is used, which normalizes for the variability due to sampling. The Monte Carlo procedure is implemented: for each child 1000 unique speech samples including 100 units (100 word tokens, 100 consonant tokens) are randomly drawn from the entire speech session. If a consonant is encountered in 95% of the 1000 speech samples, it is included in the inventory.

The consonant inventories established by the different inventory procedures were compared. The unit of analysis used in the procedure that led to the most reliable and the most accurate inventories was considered most appropriate for constructing and comparing consonant inventories from spontaneous speech data.

(C) Determine the required sample size to construct consonant inventories reliably

Once the most appropriate unit of speech selection was determined (in B), the sample size that is minimally required to reliably establish consonant inventories was at stake. The

required sample size is related to the incidence of the consonants. The more frequently a child produces a particular consonant, the higher the chance that the consonant is encountered during an observation session. This implies that only small speech samples are needed to capture highly frequent consonants, whereas large speech samples are needed to capture low frequent consonants. For example, a Dutch-speaking child frequently produces word forms starting with a /b/, such as baby ‘baby’, beer ‘bear’, bed ‘bed’, bad ‘bath’, banaan ‘banana’, whereas word forms starting with a /ʃ/, such as sjaal ‘shawl’, choco ‘chocolate’ and shampoo ‘shampoo’, are produced far less frequently. If a sample consists of, for instance, 20 WI consonants, /b/ will be more likely among them than /ʃ/. If a sample of 200 consonant tokens is taken, the chance increases that /ʃ/ is also included in the sample. This means that dependent on the size of the speech sample, only highly frequent or also low frequent consonants can be sampled and thus included in the inventory.

The granularity of the consonant inventory measurement refers to the level of frequency of the consonants included in the inventory. The granularity is low when only consonants with a high incidence can be expected (on statistical grounds) in the inventory, whereas the granularity is high when consonants with a low incidence are expected in the inventory as well. The relation between the granularity of the consonant inventory measurement and the size N of the speech sample is captured in (2).

$$(2) \quad N = \frac{\ln(1-C)}{\ln(1-p)}$$

Formula (2) is a statistical formula that is based on the hit rate method introduced in child language studies by Tomasello & Stahl (2004). It computes the minimum sample size N that is needed to capture at least one consonant. If the incidence p of a specific consonant is known, the required sample size N to draw a consonant with a confidence level C can be

computed. For instance, if the incidence of a consonant is 10%, the size of the speech sample should be at least 29 consonant tokens to capture the consonant with a confidence level of 95%.

In (C), the incidence p of a consonant in the spontaneous speech of two-year-olds was determined by means of the bootstrapping procedure: for each child and each WI and WF consonant, the mean incidence was computed over 1000 unique speech samples (with $N = 100$ consonant tokens) that were randomly drawn from the entire speech session. Then, the required sample size to include consonants in the inventory was computed for various levels of frequency – i.e. different levels of granularity of the inventory measurement.

(D) A new consonant inventory procedure

Based on the findings in (A), (B) and (C), a new procedure is introduced to construct consonant inventories in a reliable way. The procedure is then applied to the speech data of 30 Dutch-speaking two-year-olds.

Results

(A) Determine the exact relation between the amount of speech data and the size of the consonant inventory

In order to investigate whether the amount of speech material affects the size of the consonant inventories, WICIs and WFCIs were established from speech samples with varying size (range: 20 to 375 word tokens). In figures 2 and 3 respectively, the size of the WICIs and the WFCIs is plotted relative to the number of word tokens included in the sample. The maximum number of word tokens in the samples varies between the children, because the number of word tokens in the entire speech sessions differs from child to child,

as was shown in table 1. A separate graph is drawn for each child individually (S1 to S30 denote the 30 children). Each point in the graph represents the consonant inventory size corresponding to a given sample size. For each size 1000 unique speech samples were randomly selected. This implies that for each size 1000 points are presented. Subsequently, the best fitting regression line was drawn through the scatter plots and R^2 was calculated. In order to interpret the graphs adequately, it should be noted that in Dutch there are 21 WI consonants and 14 WF consonants.

[INSERT FIGURES 2 AND 3 AROUND HERE]

Three important results can be deduced from figures 2 and 3. First of all, the size of the consonant inventories is positively correlated with the number of word productions included in the sample. For all children an increase of the sample size results in an increase of the WICI size (range R^2 : 0.75-0.97, $p < 0.01$) and in an increase of the WFCI size (range R^2 : 0.76-0.97, $p < 0.01$).

Secondly, for the majority of the children, the best fitting regression is a logarithmic one: a steep increase in the beginning, but a more gradual slope and/or even a plateau near the end. This means that initially the inventory size increases rapidly with growing sample size, but if the sample size is large enough the inventory size reaches its highest level. At this level, increasing the sample size does not influence the consonant inventory anymore. The sample size required to reach this level varies between the children.

Thirdly, there is variation in inventory size between equally large samples drawn from the same observation session. For instance, for child S1 the WICI size derived from samples of 50 words varies between three and seven consonants. This variation is inherent to

sampling, hence indicating the need for implementing a normalization technique, such as the bootstrapping procedure, when constructing consonant inventories.

To conclude, the outcome of the present study (A) reveals that consonant inventories depend heavily on the size and the composition of the speech samples from which they are drawn. Therefore, normalizing for sampling bias is required. The next question to be addressed relates to the unit that should be used to determine the size of the speech sample: a particular time, or a particular linguistic unit? This question is answered in (B).

(B) Evaluate four units of speech selection

In (A) it was shown that the size of the selected material affects the number of segments in an inventory. Does the ‘unit of selection’ influence the size of the consonant inventory as well? In this study the influence of a number of units is investigated: a specific time span (e.g. the duration of the speech recording) or a linguistic unit such as the utterance, the word, or the segment.

‘Time’ is not an appropriate unit. Table 1 shows that speech samples of equal duration (60 minutes) may differ with respect to the number of word productions they incorporate. In (A), it was demonstrated that consonant inventories are dependent on the number of word productions from which they are derived. This implies that if the number of word tokens differs between speech samples, the resulting consonant inventories cannot be compared in a reliable way.

Besides a particular time period, several linguistic units may be chosen as the unit for compiling a sample: the utterance, the word and the consonant. Some inventory studies opt for ‘the utterance’ as unit of selection (e.g. Amayreh & Dyson, 2000). However, Molemans, Van Severen, van den Berg, Govaerts & Gillis (2010) show that the number of words differs between samples with the same number of utterances. More specifically, they selected

random samples of one hundred utterances from 20 minutes of speech from 30 Dutch-speaking children, and they reported considerable variation in the number of word tokens between the least and the most talkative child. Because results in (A) show that such differences lead to differences in consonant inventory size, it can be safely concluded that ‘the utterance’ is not an appropriate unit of selection.

Two units remain to be investigated: ‘the word’ and ‘the consonant’. In order to assess this issue, separate inventories were constructed for samples of 100 consonant tokens and 100 word tokens. In figure 4, box plots represent the median and the distribution of the WICI and WFCI sizes measured over all children. WICIs were established for all 30 children. WFCIs were determined for 24 children, because six children did not produce a sufficient number of WF consonants.

[INSERT FIGURE 4 AROUND HERE]

Three important observations can be made from figure 4. Firstly, the choice for a particular unit of selection influences the size of the consonant inventories. When 100 words are selected, the median WICI consists of 8.5 segments (range: 6-11), and the median WFCI consists of 4 segments (range: 0-8). When 100 consonants are selected, considerably larger WICIs and WFCIs occur: a median of 9.5 segments in the WICIs (range: 6-13) and a median of 9.5 segments in the WFCIs (range: 6-12). These differences are statistically significant in WI position (Wilcoxon Signed Rank Test: $Z=-3.898$; $p<0.001$; $n=30$) as well as in WF position (Wilcoxon Signed Rank Test: $Z=-4.301$; $p<0.001$; $n=24$), implying that consonant inventories are smaller and thus underestimated if the word is the unit of selection. Hence, the consonant appears to be the most appropriate unit of speech selection. The amount of material exerts a decisive influence: when 100 consonants are selected, the sample to be

analysed consists of 100 consonant tokens for all children. However, when 100 words are selected, the number of consonant tokens is considerably smaller: between 61.6 and 85.0 WI consonant tokens and between 8.1 and 47.0 WF consonant tokens. Thus, when selecting the word as the unit of analysis the lower number of WI and WF consonant tokens is due to the fact that the Dutch words attempted by the children do not always contain a WI and/or a WF consonant, but also to the fact that Dutch-speaking two-year-olds frequently delete WI and/or WF consonants. A median percentage of 81% of the target words contain a WI consonant, while only 63% have a WF consonant. When a word with a WI or a WF consonant is attempted, two-year-olds delete 10% of the WI consonant tokens and 37% of the WF consonant tokens. As shown in (A), differences in the amount of speech material lead to differences in the consonant inventory size.

Secondly, figure 4 demonstrates that the relation between WICIs and WFCIs depends on the unit of selection that is chosen. If words are chosen as basic units, WFCIs are significantly smaller than WICIs (Wilcoxon Signed Rank Test: $Z=-4.741$; $p<0.001$; $n=24$). However, if equal numbers of consonants are selected, WICIs and WFCIs are not significantly different in size (Wilcoxon Signed Rank Test: $Z=-1.270$; $p=0.204$; $n=24$). Position specific differences in the amount of data may explain these findings: if the word is the unit of speech selection, the number of consonant tokens is larger in WI than in WF position, implying larger inventories in WI than in WF position (cf. A). Accordingly, if the consonant is the unit of analysis, i.e. if the same number of consonant tokens is selected for both word positions, WICIs and WFCIs are equally large.

Thirdly, figure 4 shows the variation in inventory size of two-year-olds acquiring Dutch. The difference between the smallest and the largest inventory is 7 consonants in WI position and 6 consonants in WF position. Thus, even though samples with an equal number of consonants are used and a normalisation procedure is implemented, individual differences

in inventory size are still observed. This implies that other factors beside the sample size and the sampling unit influence the consonant inventories of young children.

In sum, in (A) it was shown that in order to establish consonant inventories in a reliable way, inventories should be drawn from speech samples of equal size. In (B), it was revealed that the most appropriate unit of speech selection is the consonant. This implies that consonant inventories should be drawn from samples with an equal number of consonant tokens. But how many consonant tokens are needed to reliably measure consonant inventories? This question will be addressed in (C).

(C) Determine the required sample size to construct consonant inventories reliably

The present study aims to discover the number of consonants that are minimally required to reliably establish consonant inventories. The required sample size is dependent on the granularity of the consonant inventory measurement, in the sense that consonants with a high incidence will already show up in consonant inventories drawn from relatively small samples, while consonants with a relatively low incidence require much larger samples. In order to study the exact relationship between the size of the speech sample and the granularity of the consonant inventory measurement the incidence of all Dutch consonants in the speech corpus of the 30 toddlers acquiring Dutch is determined.

Figures 5 and 6 demonstrate the median incidence of respectively WI and WF consonants computed over all 30 children. In WF position, only data of 24 children were included in the analysis, because six children did not produce sufficient WF consonants. The median incidence of WI consonants ranges between <0.1% and 13%. The median incidence of WF consonants ranges between 0.8% and 15%. Based on their incidence, consonants are classified in four groups: high-frequent ($\geq 10\%$), mid-frequent (5-10%), low-frequent (2-5%), and rare consonants (0.1-2%). More specifically, the WI consonants /p/, /j/, /t/ and the WF

consonants /t/, /n/, /s/, /k/ are categorised as high-frequent. The WI consonants /n/, /k/, /d/, /b/, /m/, and the WF consonants /m/, /p/ are mid-frequent. The WI consonants /w/, /h/, and the WF consonants /χ/, /j/, /w/, /r/ are classified as low-frequent. The WI consonants /l/, /z/, /s/, /g/, /ʃ/, /v/, /r/, /χ/ and the WF consonants /ʃ/, /f/, /l/, /ŋ/ are rare. The median frequency of the WI consonants /f/, /r/, /χ/, /ʒ/ is lower than 0.1%.

[INSERT FIGURES 5 AND 6 AROUND HERE]

The granularity of the inventory is related to the size of the speech sample. The larger the speech sample, the higher the granularity: in large samples even rare consonants show up. Figure 7 gives an overview of the sample size needed to draw at least one consonant with an incidence p . The required sample size increases logarithmically when the incidence p decreases (or, in other words: when the granularity increases).

The minimal sample size N that is needed to capture at least one high-, mid-, or low-frequent and rare consonant with a confidence level of 95% is determined for the four frequency groups. Applying formula (2), the sample size N required to select high-frequent, mid-frequent and low-frequent consonants reliably is respectively 29, 59 and 149 consonant tokens. Rare consonants have an incidence between 0.1% and 2% in WI position and between 0.8% and 2% in WF position. The number of consonant tokens required for sampling rare consonants with an incidence of 0.1% and 0.8% is very high: 2995 and 373 tokens.

[INSERT FIGURE 7 AROUND HERE]

Based on these findings, it can be concluded that a speech sample of at least 2995 WI and 373 WF consonant tokens is needed to discover the complete inventory of a two-year-old child acquiring Dutch. On top of that, the inventory measurement should control for variation caused by sampling bias. Because no such procedure exists yet, a new procedure is proposed in (D).

(D) A new consonant inventory procedure

The findings in (A), (B) and (C) indicate the need for a new procedure that reliably constructs and compares consonant inventories. Ideally, WICIs and WFCIs should be derived from speech samples with at least 2995 WI consonant tokens and 373 WF consonant tokens. In order to normalize for variation due to sampling a bootstrapping procedure should be implemented, similar to the one outlined in the flowchart in figure 1, but this time with a predetermined sample size, instead of an increasing sample size. A consonant occurring in at least 95% of the inventories drawn after each random sampling round, can be included in the inventory with 95% confidence.

For the sake of clarity, the procedure is briefly summarized in the flowchart in figure 8.

[INSERT FIGURE 8 AROUND HERE]

Two variables need to be further specified in the flowchart in figure 8: the number of random sampling rounds to be executed (variable i), and the number of consonants required in the speech recording (variable K). Given that $N=2995$, the required number of WI consonant tokens, either variable K or variable i need to be specified. In this study the value of variable i was arbitrarily set to 1000. This means that 1000 unique random samples were drawn from

the material. How many WI consonant tokens (K) should be available in order to draw 1000 unique combinations of 2995 WI consonants? For the required size $N=2995$ consonant tokens, formula (1) computes the size of K : minimally 2996 WI consonant tokens are needed to construct 1,000 unique combinations of 2995 WI consonant tokens. For WFCIs, K should be at least 375 WF consonant tokens in order to establish 1000 unique combinations of 373 WF consonants. Table 1 demonstrates that most children in the speech corpus in the present study did not produce that many WI and/or WF consonant tokens, i.e. the required sample size N to construct the complete inventory. This implies that the granularity of the consonant inventory measure has to decrease.

The speech corpus of the 30 Dutch-speaking children participating in the present study is evaluated with respect to various granularity levels of the consonant inventories. The bottom panel of table 2 presents the number of children that have sufficient consonant tokens for the construction of inventories with varying levels of granularity. Results are reported separately for WI and WF position. In both word positions, the data of all children contain more than 32 consonant tokens. This implies that for all 30 children inventories of high-frequent consonants can be drawn reliably in WI as well as in WF position. In WI position, inventories incorporating mid- and low-frequent consonants can be reliably established for all 30 children, while in WF position, inventories of mid-frequent and low-frequent consonants cannot be reliably constructed for all 30 children. Table 2 shows that these inventories can only be reliably determined for 26 and 20 children respectively, due to limitations in the available number of WF consonant tokens.

[INSERT TABLE 2 AROUND HERE]

Discussion

Widespread phonological measures, such as consonant inventories, are often drawn from speech samples that vary considerably in size. Because it is unclear how the size and the nature of the speech data influence the resulting consonant inventories, the present study examines the effect of these two factors. A spontaneous speech corpus of 30 typically developing two-year-olds acquiring Dutch is used to (A) determine the exact relation between the amount of speech and the size of the inventory; (B) evaluate various units of speech selection (amount of time versus linguistic units) and (C) specify the required sample size for the construction of consonant inventories. In study (D), a new procedure is proposed that takes the findings in (A), (B) and (C) into account.

(A) Determine the exact relation between the amount of speech data and the size of the consonant inventory.

In the present study it is shown that the size of the speech sample is strongly correlated with the consonant inventory drawn from it: the larger the sample, the larger the inventory. Notwithstanding this obvious link between the sample size and the inventory size, the procedure that is commonly applied in prior inventory studies (i.a. Stoel-Gammon, 1985; 1987; Kunnari, 2003; Morris, 2009) does not control for sample size. Consequently, the reported inventories depend heavily on the size of the speech samples from which they are drawn, and therefore may not be sufficiently reliable.

The speech selection and consonant inclusion criterion used in this widespread procedure may explain how the variation in sample size arises and how this variation affects the consonant inventories. The speech selection criterion prescribes that all intelligible words in the speech recordings are included in the analysis. This implies that due to differences in the volubility of the children, the speech recordings differ in the number of (intelligible) words

(cf. Molemans et al., 2010). Even if the selection is restricted to only two variants (tokens) of a particular type, samples of unequal size result from the procedure. The outcome of study A reveals that consonant inventories depend heavily on the size of the speech samples from which they are drawn. Thus, consonant inventories that are constructed from speech samples established according to the commonly applied speech selection criterion are not sufficiently reliable, unless a consonant inclusion criterion is used that compensates for the variation in the sample sizes¹.

However, the consonant inclusion criterion used in Stoel-Gammon (1985; 1987), Kunnari (2003) and Morris (2009) prescribes that only consonants that occur in at least two different word types can be added to the inventory regardless of the size of the speech selection. As a result, large speech samples have an advantage over small ones: the more word tokens, the more word types can be incorporated in the sample, the higher the chance that at least two word types share the same consonant. Therefore, it is important to draw consonant inventories from speech samples with equal size to make reliable comparisons between children or across time. But how should the size of the speech selection be specified? More specifically, what unit of speech selection should be used?

(B) Evaluate four units of speech selection.

In study (B) four units are assessed: a particular time period and three linguistic units, e.g. the utterance, the word or the segment. The present study and Molemans et al. (2010) show that in speech samples with a fixed time period or a fixed number of utterances, the number of word tokens varies considerably. Because the number of word tokens is highly correlated with the size of the consonant inventory (cf. A), it can be concluded that ‘the time duration’ or ‘the utterance’ are not the appropriate units.

Moreover, it turns out that speech sessions with a fixed number of words vary substantially in the number of WI and WF consonant tokens. Because the size of the speech data is strongly related to the size of the consonant inventory, the word is not the best sampling unit either. In addition, more accurate inventories are constructed when the sampling unit is the segment rather than the word, implying that the word is a less convenient unit of analysis than the segment.

To summarize, these findings demonstrate that ‘the consonant’ is most appropriate to construct consonant inventories, hence indicating that consonant inventories should be drawn from speech samples that incorporate equal numbers of consonant tokens. The next question addressed is: how many consonant tokens are needed to construct the complete consonant inventory?

(C) Determine the required sample size to construct consonant inventories reliably

Because the inventory of the target language is restricted in the number of consonantal types included, the number of consonant tokens that is needed to capture the complete inventory should be limited as well. More specifically, the relation between the size of the consonant inventory and the size of the speech sample is logarithmic. At first a steep increase in the inventory size is observed, giving way to a more gradual slope and eventually a plateau is reached. This implies that for a particular sample size N , increasing the size of the speech data does not influence the size of the consonant inventory anymore. In (C), this crucial sample size N is computed.

The number of consonant tokens required for sampling all consonantal types depends on their incidence: only a small set of consonant tokens suffices to capture highly frequent consonantal types, while a large number of tokens is needed to sample consonants that are rarely used by the child. More specifically, the required sample size N to draw all WI and WF

consonants with a confidence level of 95% is 2995 WI consonant tokens and 373 WF consonant tokens.

However, the present speech data do not consist of that many consonant tokens for each child individually. CHILDES corpora (MacWhinney, 2000) are not dense enough either to support a consonant inventory measure that captures the complete inventory (Tomasello & Stahl, 2004). How, then, can the consonant inventory be measured in corpora that are not large enough?

(D) A new inventory procedure

In study (D), methodological guidelines are provided for reliably constructing and comparing inventories from spontaneous speech productions of children: select samples with an equal size (cf. A) and choose the consonant as the unit of speech selection (cf. B). Notwithstanding that speech samples with an equal number of consonants are selected, the resulting consonant inventories differ in size – even when these speech samples are randomly selected from the same child and the same recording session (cf. A). To control for this sampling bias, it is advised to implement a normalisation procedure, such as the Monte Carlo procedure (Efron, 1979; Baayen, 2008). Concretely, not one but a large number of samples (e.g. 1000) with a fixed number of consonant tokens N should be randomly selected from the entire speech recording. A consonant that occurs in 95% of the inventories constructed after each sampling round, can be added to the inventory with 95% confidence (see figure 8).

Ideally, the samples should contain $N=2996$ WI consonant tokens and $N=375$ WF consonant tokens to be able to establish the complete inventory. However, the existing children's speech corpora do not comprise that many consonant tokens. The present study proposes an alternative. Equally large samples have to be selected from all speech recordings to allow for reliable comparisons between children and across time. This implies that the size

of the speech recording of the least talkative child is the largest potential sample size that can be applied in the selection of speech material for all children in the corpus. This size also determines the level of granularity of the consonant inventory measure — i.e. the smallest potential incidence of the consonantal types that can be in the inventory. For example, if the speech session of the least voluble children incorporates 32 consonant tokens, the largest potential sample size is 32 consonant tokens for all children in the corpus. This implies that for all children only consonantal types with an incidence of at least 10% are included in the inventory. If speech samples with more consonant tokens are selected, consonant inventories cannot be constructed for the least voluble children in a reliable way. For the purpose of replication inventory studies should report the exact granularity of the inventory, and the sample size N that is used.

The granularity of the inventory measurement varies across languages. The granularity, and thus the required number of consonant tokens to sample the complete inventory shall be smaller for children acquiring a language with less consonantal types. For instance, only 11 WI consonants are used in Finnish. Provided that they have an equal incidence, smaller speech samples can be used to arrive at the same precision as for Dutch.

Finally, the proposed methodological guidelines to construct consonant inventories can be extended to other inventories, such as vowel or syllable inventories. For these inventories, the unit of speech selection should respectively be the vowel or the syllable. This implies that the samples that are compared should contain equal numbers of vowels or syllables. Note that the minimally required sample size depends on the incidence of the vowels or the syllables, and most likely differs from the one needed for the consonant inventory.

Theoretical and clinical implications

The implementation of these methodological guidelines leads to new and more reliable insights in the process of phonological acquisition. Firstly, the number of consonant ‘tokens’ that are produced by two-year-olds acquiring Dutch differs between word positions (WI>WF), but the number of consonantal ‘types’ does not differ significantly between the two positions. According to the phonology of (adult) Dutch, more different consonantal types can occur in WI (n=21 consonants) than in WF position (n=14 consonants). This implies that – if the children’s inventories are compared with the target inventories – the children’s WFCI’s (the median size = 9/14) are relatively larger than their WICI’s (the median size = 9/21). It can be speculated that it is easier to acquire consonantal types if there are less consonantal types in the target language. For instance, in languages with more consonantal types in WF position (as in English), it should be harder to acquire WF consonants (cf. Ingram, 1981; Stoel-Gammon, 1985, 1987, 2002).

Secondly, the current investigation reports less individual variation with respect to the size of the consonant inventories in Dutch as compared to other languages. For Dutch-speaking children (age range: 1;10–2;0), the range amounts to seven consonants in WI position and six in WF position. For two-year-olds acquiring English, Stoel-Gammon (2002) finds that the difference between the largest and the smallest consonant inventory amounts to 10 consonants (the target language allows 24 consonants in both word positions). Savinainen-Makkonen et al. (2008) report ranges from 0 to 12 different WI and from 0 to 5 WF consonants in the inventories of two-year-olds acquiring Finnish (the target language allows 11 WI and 5 WF consonants). The large variation reported in the latter studies may be explained partly by the differences between the consonant inventories in the target languages. But part of the explanation may also be found in methodological differences: prior inventory studies neither analysed speech samples with an equal number of consonant tokens, nor did they use a bootstrapping procedure to control for sampling bias.

The findings in the present study have also important clinical implications. If researchers or clinicians want to compare consonant inventories between children and across time, they should select speech material that is equal in size in order to make reliable comparisons. If speech samples of different size are taken, the observed individual variation in consonant inventory size may be –to some extent- caused by differences in the amount of data. For instance, it has been reported that late talkers are (far) less voluble than typically developing children (Paul & Jennings, 1992; Rescorla & Ratner, 1996). Hence, typically developing children talk more than late talkers in the same amount of time, implying that less data are available for the latter group. In addition, late talkers were found to have smaller consonant inventories than typically developing children (e.g. Paul & Jennings, 1992; Thal et al., 1995; Rescorla & Ratner, 1996). In the present study significant correlations were revealed between the amount of data analysed and the size of the consonant inventories, hence the smaller inventories of the late talkers may well be attributable to the smaller speech samples that were analysed of the less voluble late talkers. It would be interesting to investigate if late talkers still have smaller inventories than typically developing children when the amount of speech material is controlled.

Limitations of the proposed procedure

The present paper introduces a new procedure to reliably construct consonant inventories using spontaneous speech from young children. Although this procedure normalizes the differences in the size of the selected speech samples and reduces the effect of sampling bias, this procedure has some restrictions too.

A first issue concerns the type of speech material. Consonant inventories can be drawn from elicited speech or from spontaneous speech. An important advantage of elicitation tasks is that the target productions can be controlled: each sound can be evoked in

all possible word positions in a fixed number of opportunities, so that the sample is balanced across children and even the rarer consonants can be sampled sufficiently. The disadvantages are that a large attention span, a relatively large vocabulary, and a good understanding of the task are required. Because these preconditions are not necessarily met in children below age 2;0, spontaneous speech is more appropriate than elicited speech to investigate language productions of very young infants and toddlers.

However, collecting and transcribing spontaneous speech, especially from young children, is very time-consuming. This may be an important obstacle in clinical practice. Moreover, human transcriptions are to a certain extent subjective. This implies that -- especially for transcriptions of spontaneous speech of young children -- two transcribers may disagree on the identity of the consonants, and this may have an effect on the size of the resulting consonant inventory.

A supplementary analysis of the transcriptions that resulted from the reliability check in the present study reveals close, but not complete, agreement between the primary and the second transcribers (interrater agreement). In approximately 40% of the transcriptions complete agreement was reached on the size of WICIs and WFCIs. Higher agreement scores were observed for the intrarater agreement: complete agreement was accomplished in 80% of the speech samples. If the inventories had a different size, they only diverged in one or sporadically two consonants. Consequently, when constructing consonant inventories in clinical or scientific practice, such inaccuracies should be taken into consideration.

A second issue pertains to the limitations in the amount of speech material available in speech samples of particular individuals or groups. For example, very young children or children with phonological disorders (occasionally) do not produce a large enough sample. Or, even if they are sufficiently voluble, they may not produce many WI or WF consonant tokens, because they delete WI and/or WF consonants. Then, the clinician or researcher has

three possibilities: (1) either collect more data, or (2) in case of a group study, exclude children with too small speech samples from the analysis. (3) If speech samples of too few of the children contain ample consonant tokens, reliably constructing consonant inventories for this particular group of children is impossible. In the latter instance and if the children have sufficiently large attention span, a relatively large vocabulary, and a good understanding of the task, an elicitation task can be used, such as a picture-naming task (e.g., Gierut, 2008).

A third issue relates to the analysis of the data. First of all, the consonant inventory procedure introduced in this paper cannot be applied manually, but it requires computational resources to conduct the ‘bootstrapping procedure’. Secondly, the proposed procedure does not incorporate any relational criteria. The techniques used in this study may be helpful to develop a consonant inventory procedure using a relational analysis. In that way, the relationships between the actually produced sounds and the target sounds as well as the relationships between the sounds produced by the child — e.g. two sounds may be in complementary distribution and act as two variants of a single adult phoneme — are addressed. In a similar vein, the proposed procedure applies a simple inclusion criterion. It can be extended using a tool that yields the frequency at which different sounds are produced.

Conclusion

The present study investigated the reliability of a widespread measure of children’s spontaneous speech, namely the consonant inventory. It was shown that the size of the selected speech sample as well as the unit of speech selection (e.g. a particular amount of time, a number of utterances, words or segments) strongly influence the resulting consonant inventories. Based on these study outcomes, methodological guidelines are proposed that control both factors, and thus increase the reliability of the inventory. Implementing these guidelines provides the opportunity to reliably assess the regularities and the variation present

in a normal population, and to reliably compare clinical groups with typically developing children.

Footnotes

(1) Ingram (1981; 1989) and Carson, et al. (2003) used a criterion that aims to take the differences in sample size into account. The consonant inclusion criterion prescribes that a consonant has to appear with an incidence of at least $\frac{1}{2}$ of the frequency criterion (see formula 1.1).

$$(1.1.) \text{ Frequency criterion} = \frac{\text{All word types}/\text{All unique word forms}}{50}$$

This implies that a consonant should occur more often in a large sample than in a small sample in order to be included in the inventory. Note that the unit of analysis is the word, and not the consonant. In study B, it was shown that the consonant is a more appropriate unit than the word.

Acknowledgments

The research reported in this paper was supported by a TOP-BOF grant of the Research Council of the University of Antwerp. Thanks are due to the editor and two anonymous reviewers for insightful comments. We are especially indebted to all of the children and parents who participated in this study.

Declaration of Interest

The authors report no conflicts of interest.

References

- Amayreh, M., & Dyson, A. (2000). The acquisition of Arabic consonants. *Journal of Speech, Language and Hearing Research, 41*, 642-653.
- Baayen, H. (2008). *Analysing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Beers, M. (1995). *The phonology of normally developing and language-impaired children*. PHD Thesis, University of Amsterdam, Amsterdam.
- Booij, G. (1995). *The phonology of Dutch*. Oxford: Clarendon Press.
- Bornstein, M., Painter, K., & Park, J. (2002). Naturalistic language sampling in typically developing children. *Journal of Child Language, 29*, 687-699.
- Carson, C., Klee, T., Carson, D., & Hime, L. (2003). Phonological profiles of 2-year-olds with delayed language development: predicting clinical outcomes at age 3. *American Journal of Speech-Language Pathology, 12*, 28-39.
- Catano, L., Barlow, J., & Moyna, M. (2009). A retrospective study of phonetic inventory complexity in acquisition of Spanish: implications for phonological universals. *Clinical Linguistics & Phonetics, 23*, 446-472.
- Dinnsen, D., Chin, S., Elbert, M., & Powell, T. (1990). Some constraints on functionally disordered phonologies: phonetic inventories and phonotactics. *Journal of Speech, Language and Hearing Research, 33*, 28-37.
- Dyson, A. (1988). Phonetic inventories of 2- and 3-year-old children. *Journal of Speech and Hearing Disorders, 53*, 89-93.
- Edwards, J., & Beckman, M. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics, 22*, 939-958.
- Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics, 7*, 1-26.

- Elffers, C., Van Bael, C. & Strik, H. (2005). ADAPT: algorithm for dynamic alignment of phonetic transcriptions (Internal report). Retrieved from Radboud University Nijmegen, Department of Language & Speech website:
<http://lands.let.kun.nl/literature/elffers.2005.1.pdf>.
- Fikkert, P. (1994). *On the acquisition of prosodic structure*. PHD Thesis, Holland Academic Press, The Hague.
- Gierut, J. (2008). Fundamentals of experimental design and treatment. In D. Dinnsen & J. Gierut (eds.), *Optimality Theory, phonological acquisition and disorders*. (pp. 93-118). London: Equinox Publishing Ltd.
- Ingram, D. (1981). *Procedures for the phonological analysis of children's language*. Baltimore: University Park Press.
- Kunnari, S. (2003). Consonant inventories: a longitudinal study of Finnish-speaking children. *Journal of Multilingual Communication Disorders, 1*, 124-131.
- MacWhinney, B. (2000). *The CHILDES project: tools for analysing talk: transcription format and programs*. Mahwah: Lawrence Erlbaum Associates.
- McKee, G., Malvern, D., & Richards, B. (2000). Measuring vocabulary diversity using dedicated software. *Literary and Linguistic Computing, 15*, 323-337.
- Molemans, I., Van Severen, L., van den Berg, R., Govaerts, P., & Gillis, S. (2010). Spraakzaamheid van Nederlandstalige baby's en peuters. *Logopedie, 23*, 12-23.
- Morris, S. (2009). Test- retest reliability of independent measures of phonology in the assessment of toddlers' speech. *Language, Speech, and Hearing Services in Schools, 40*, 46-52.
- Paul, R., & Jennings, P. (1992). Phonological behaviour in toddlers with slow expressive language development. *Journal of Speech and Hearing Research, 35*, 99-107.
- Pye, C., Ingram, D., & List, H. (1987). A comparison of initial consonant acquisition in

- English and Quiché. In K. Nelson & A. van Kleeck (Eds.), *Children's language*. (Vol. 6, pp. 175-190). Hillsdale: Erlbaum.
- Rescorla, L., & Ratner, N. (1996). Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). *Journal of Speech and Hearing Research, 39*, 153-166.
- Robb, M., & Bleile, K. (1994). Consonant inventories of young children from 8 to 25 months. *Clinical Linguistics & Phonetics, 8*, 294-320.
- Rowland, L., Fletcher, S., & Freudenthal, D. (2008). How big is enough? Assessing the reliability of data from naturalistic samples. In H. Behrens (Ed.), *Trends in language acquisition research 6: corpora in language acquisition research. History, methods, perspectives*. (Vol. 6, pp. 1-25). Amsterdam: John Benjamins Publishing Company.
- Sander, E. (1972). When are speech sounds learned? *Journal of Speech and Hearing Disorders, 37*, 55-63.
- Savinainen-Makkonen, T., Kunnari, S., & Paavola, L. (2008). *Consonant inventories of Finnish-speaking children of age two*. Paper presented at the XI international congress for the study of child language (IASCL). Edinburgh.
- Stoel-Gammon, C. (1985). Phonetic inventories, 15-24 months: a longitudinal study. *Journal of Speech and Hearing Research, 28*, 505-512.
- Stoel-Gammon, C. (1987). Phonological skills of 2-year-olds. *Language, Speech, and Hearing Services in Schools, 18*, 323-329.
- Stoel-Gammon, C. (2002). Intervocalic consonants in the speech of typically developing children: emergence and early use. *Clinical Linguistics & Phonetics, 16*, 155-168.
- Stokes, S., Klee, T., Carson, C., & Carson, D. (2005). A phonemic implicational feature hierarchy of phonological contrasts for English-speaking children. *Journal of Speech, Language and Hearing Research, 48*, 817-833.
- Stokes, S., & To, C. (2002). Feature development in Cantonese. *Clinical Linguistics &*

Phonetics, 16, 443-459.

- Taelman, H., Durieux, G., & Gillis, S. (2004). Notes on Ingram's whole-word measures for phonological development. *Journal of Child Language*, 32, 391-405.
- Templin, M. (1957). Certain language skills in children, their development and interrelationships. In Institute of Child Welfare (Ed.), *Monograph Series*. (Vol. 26). Minneapolis: University of Minnesota Press.
- Thal, D., Oroz, M., & McCaw, V. (1995). Phonological and lexical development in normal and late-talking toddlers. *Applied Psycholinguistics*, 16, 407-424.
- Tomasello, M., & Stahl, D. (2004). Sampling children's spontaneous speech: how much is enough? *Journal of Child Language*, 31, 101-121.
- Vihman, M., & McCune, L. (1994). When is a word a word? *Journal of Child Language*, 21, 517-542.
- Watson, M., & Scukanec, G. (1997). Profiling the phonological abilities of 2-year olds: a longitudinal investigation. *Child Language Teaching and Therapy*, 13, 3-14.
- Xanthos, A., & Gillis, S. (2010). Quantifying the development of inflectional diversity. *First Language*, 30, 175-198.
- Zamuner, T. (2003). *Input-based phonological acquisition*. London: Routledge.
- Zink, I., & Lejaegere, M. (2002). *N-CDI's: lijsten voor communicatieve ontwikkeling: aanpassing en hernormering van de MacArthur CDIs van Ferguson et al.* Leuven: Acco.
- Zmarich, C., & Bonifacio, S. (2005). Phonetic inventories in Italian children aged 18-27 months: a longitudinal study. *Interspeech*, 757-760.

Figure Captions

Figure 1. A flowchart of the procedure for drawing consonant inventories from speech samples with increasing size, using the technique of bootstrapping.

Figure 2. The size of the WICI as a function of the number of word tokens in the samples: the best fitting regressions.

Figure 3. The size of the WFCI as a function of the number of word tokens in the samples: the best fitting regressions.

Figure 4. The effect of the unit of selection (word versus consonant) on the size of the consonant inventory in 30 Dutch-speaking two-year-olds.

Figure 5. The median incidence of WI consonants in the spontaneous speech of 30 two-year-olds acquiring Dutch.

Figure 6. The median incidence of WF consonants in the spontaneous speech of 30 two-year-olds acquiring Dutch.

Figure 7. The required sample size (N) as a function of the incidence (p).

Figure 8. Flowchart of the procedure proposed for drawing consonant inventories from speech samples with size N , using the technique of bootstrapping.

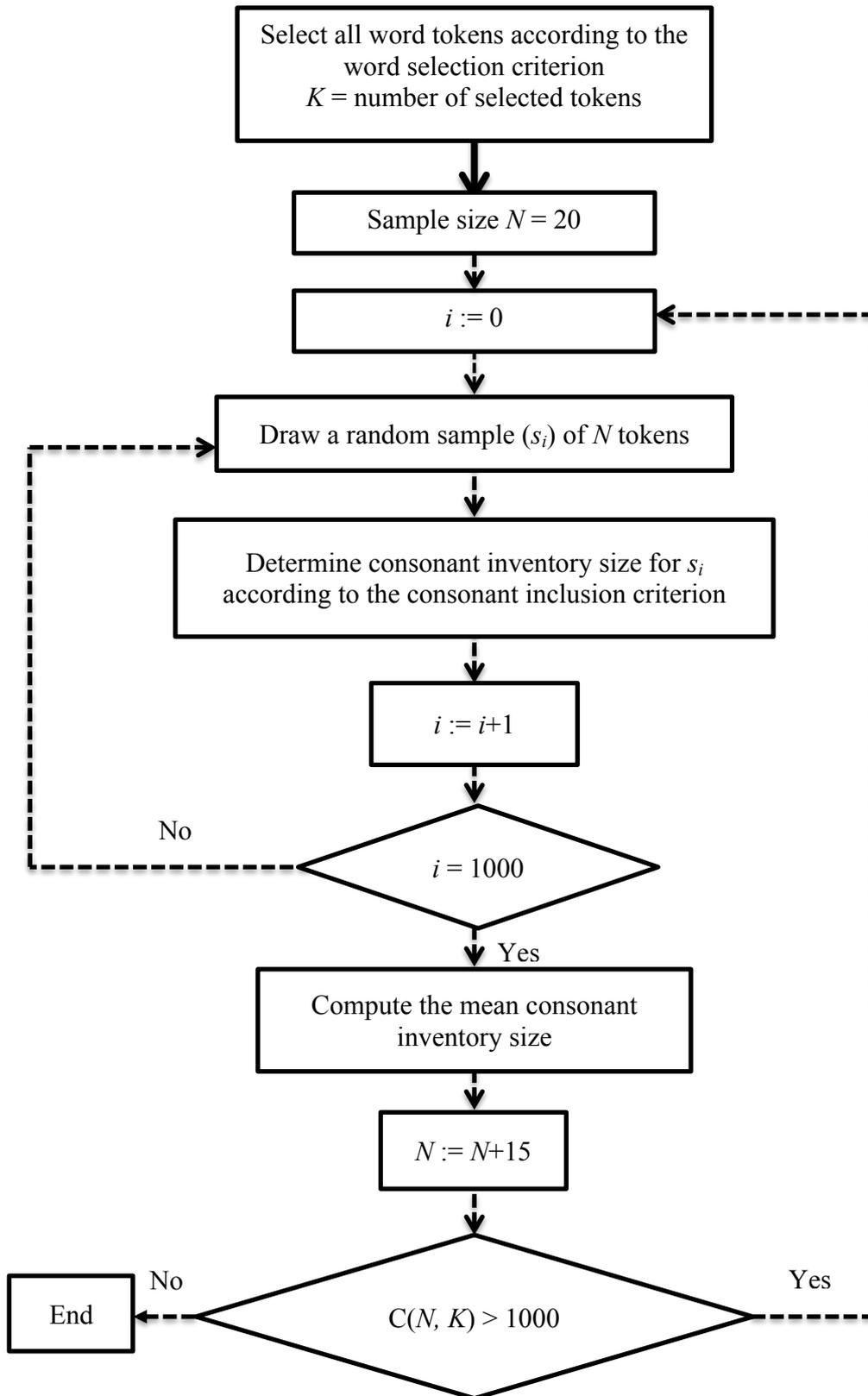


Figure 1. A flowchart of the procedure for drawing consonant inventories from speech samples with increasing size, using the technique of bootstrapping.

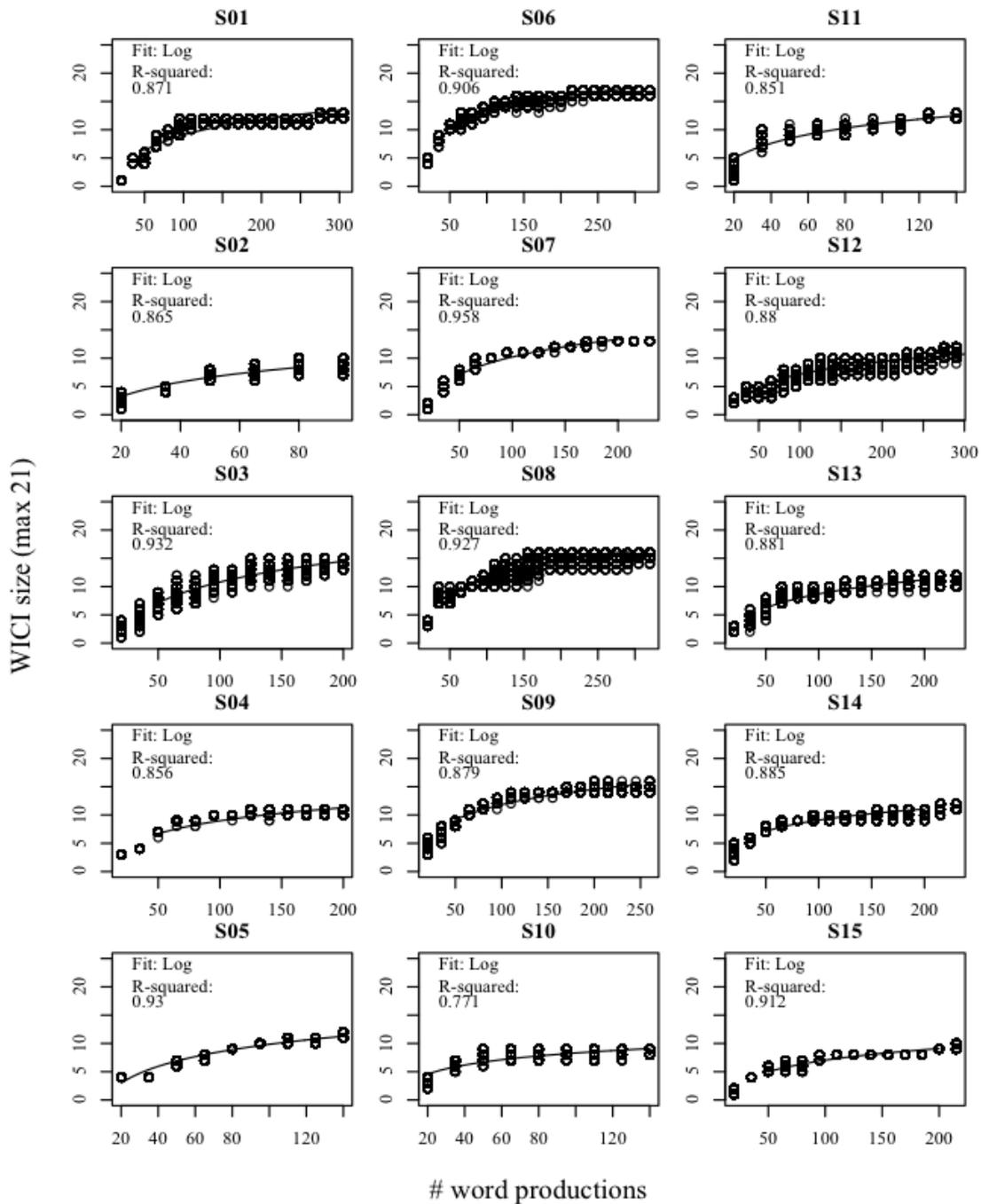


Figure 2. The size of the WICI as a function of the number of word tokens in the samples: the best fitting regressions.

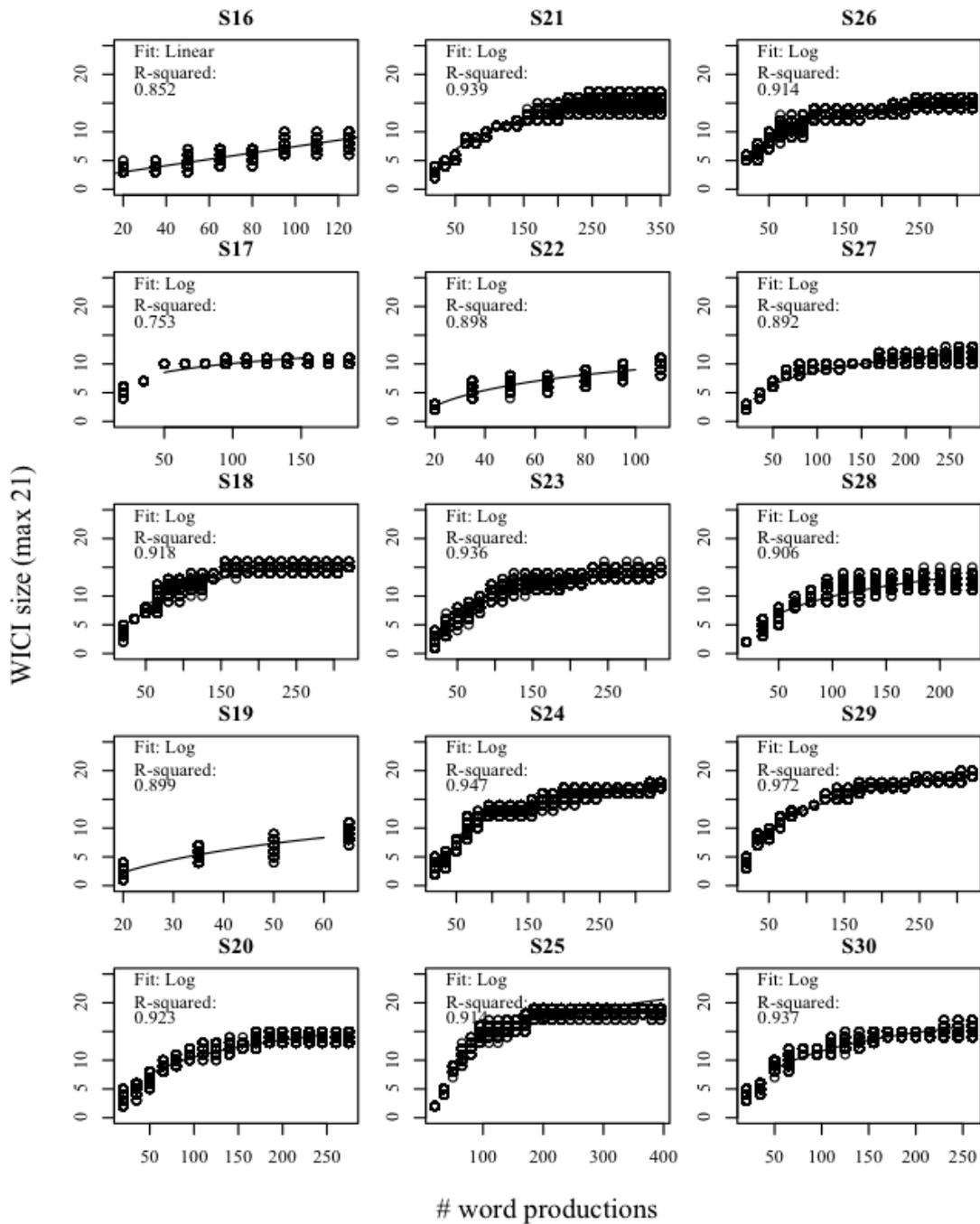


Figure 2 (continued). The size of the WICI as a function of the number of word tokens in the samples: the best fitting regressions.

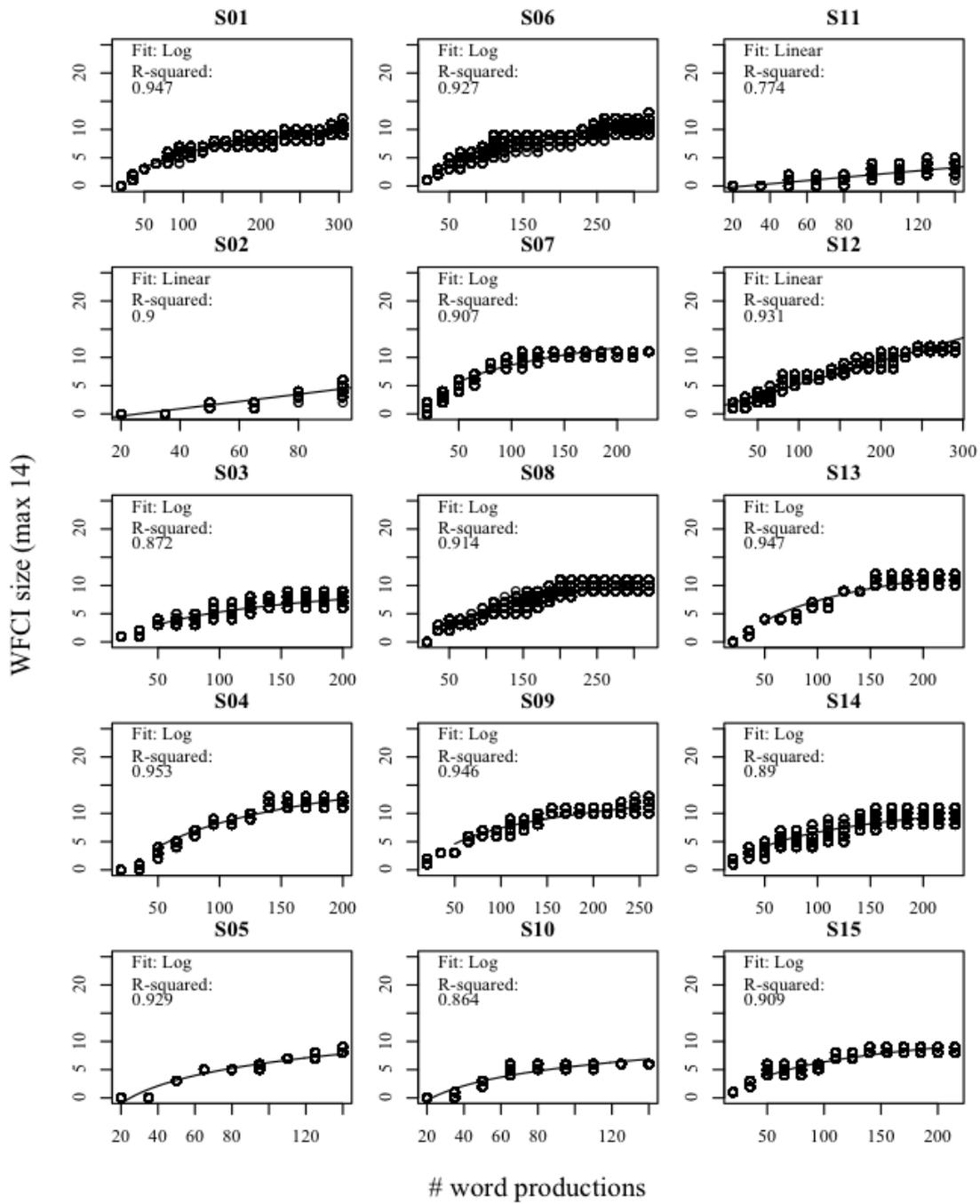


Figure 3. The size of the WFCI as a function of the number of word tokens in the samples: the best fitting regressions.

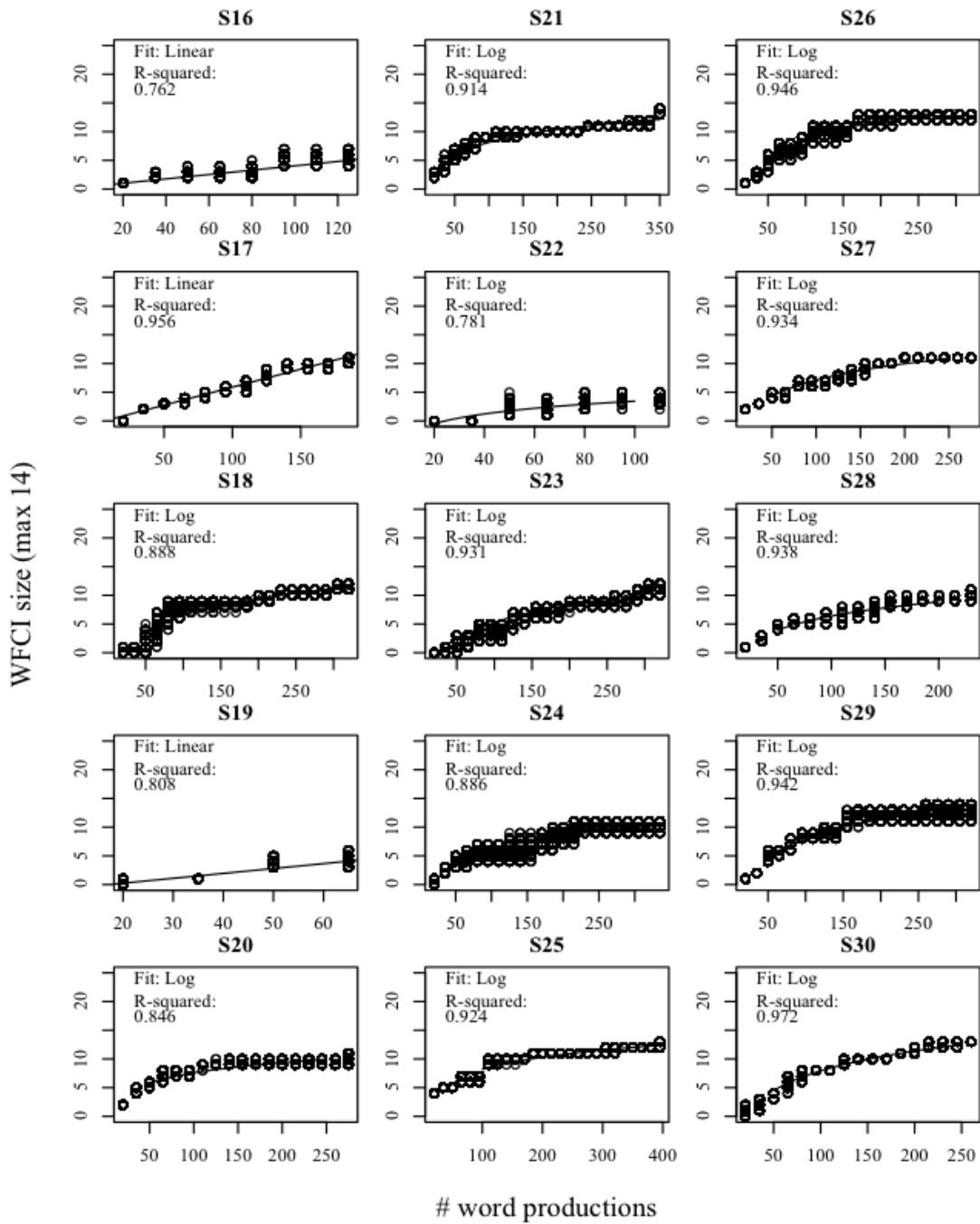


Figure 3 (continued). The size of the WFCI as a function of the number of word tokens in the samples: the best fitting regressions.

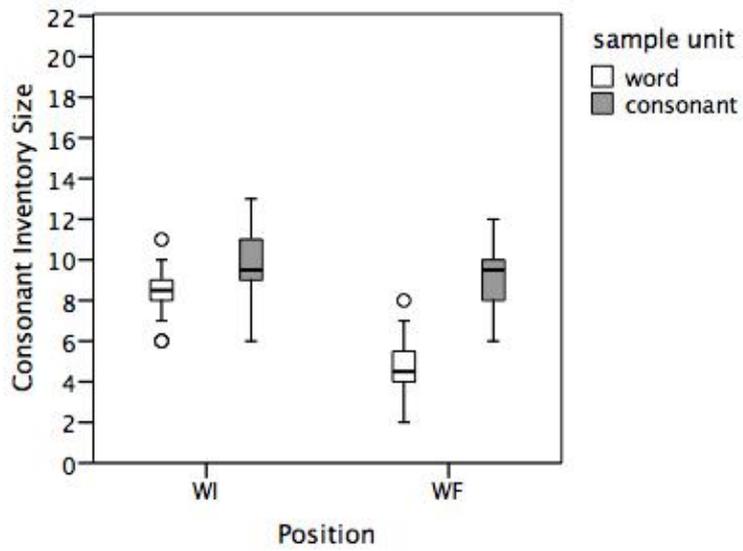


Figure 4. The effect of the unit of selection (word versus consonant) on the size of the consonant inventory in 30 Dutch-speaking two-year-olds.

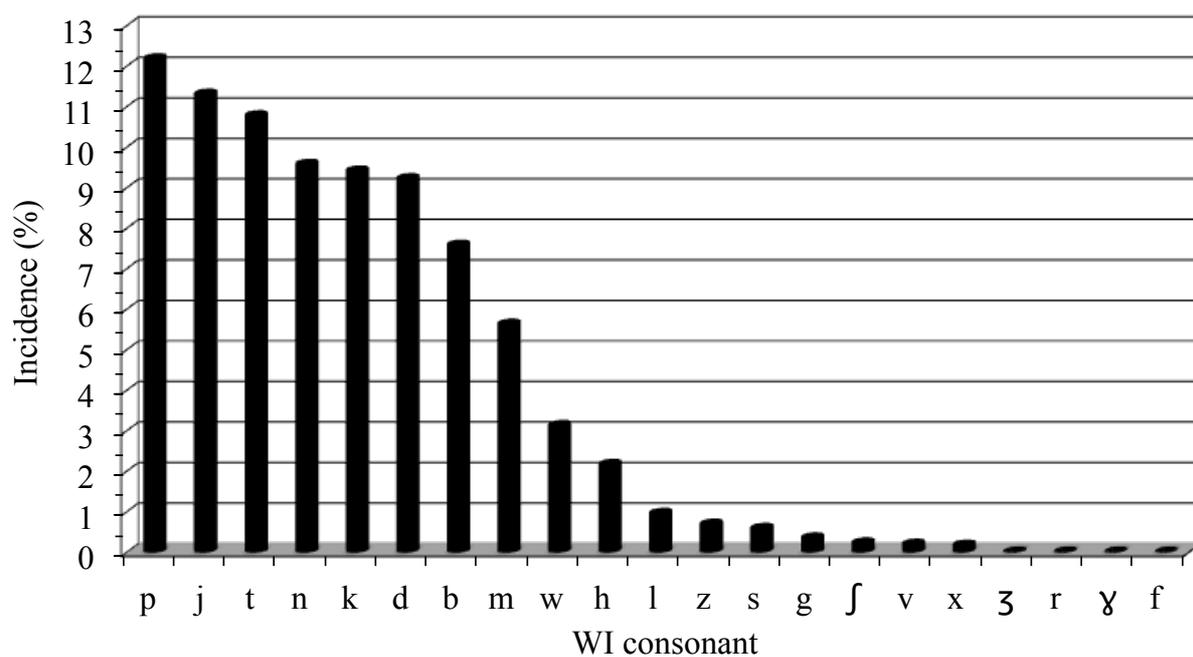


Figure 5. The median incidence of WI consonants in the spontaneous speech of 30 two-year-olds acquiring Dutch.

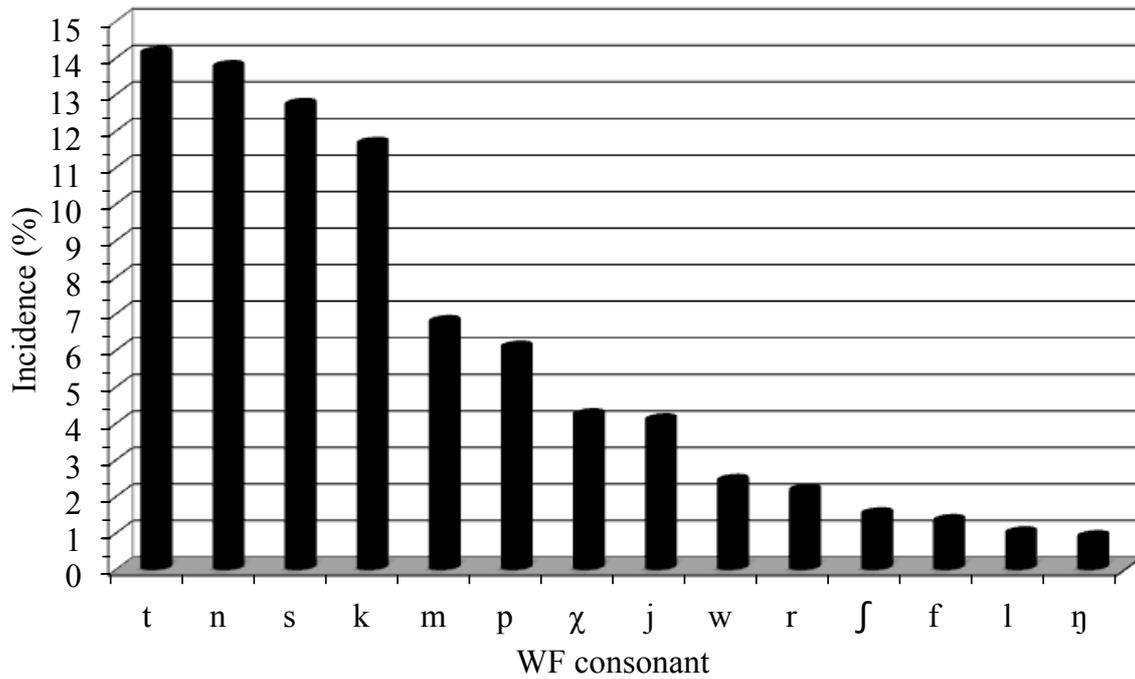


Figure 6. The median incidence of WF consonants in the spontaneous speech of 30 two-year-olds acquiring Dutch.

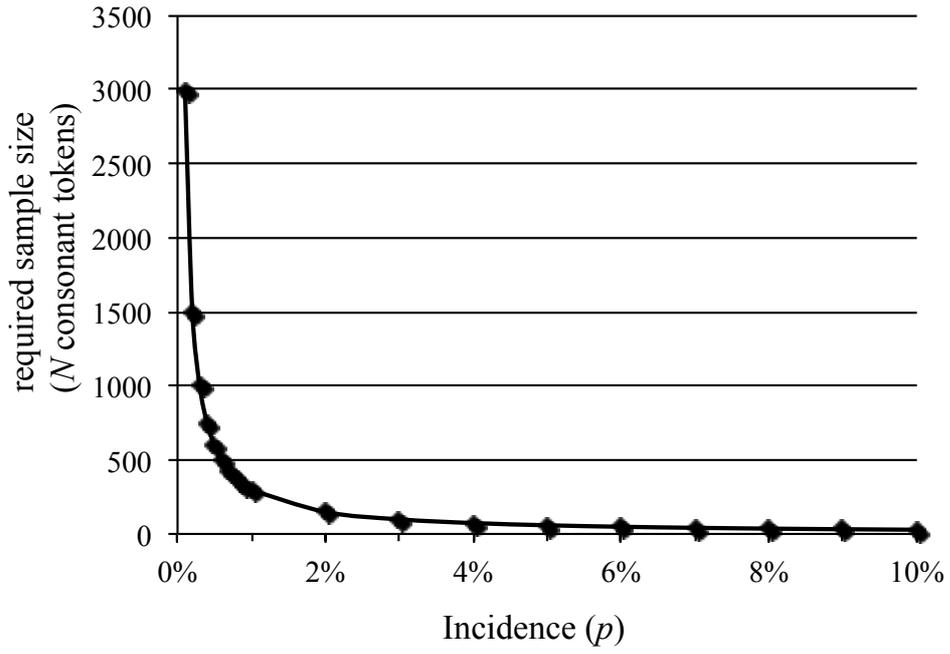


Figure 7. The required sample size (N) as a function of the incidence (p).

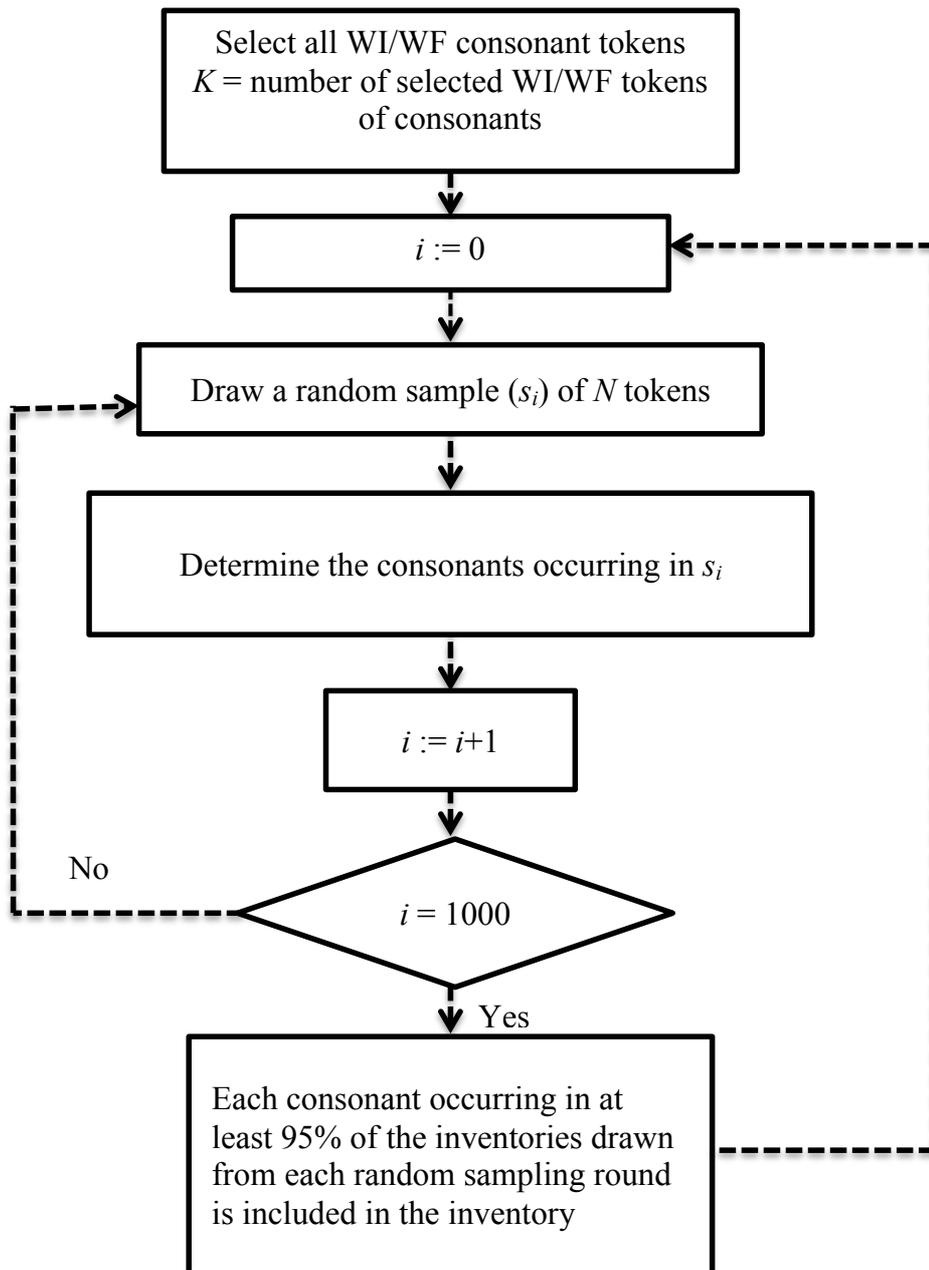


Figure 8. Flowchart of the procedure proposed for drawing consonant inventories from speech samples with size N , using the technique of bootstrapping.

Table Captions

Table 1. Individual variation in the size of the entire speech sessions.

Table 2. The granularity of the consonant inventory measurement in relation to the available speech data: group data of 30 Dutch-speaking two-year-olds.

Table 1. Individual variation in the size of the entire speech sessions.

	<i>Number of word tokens</i>	<i>Number of word types</i>	<i>Number of word- initial consonants</i>	<i>Number of word- final consonants</i>
<i>Median</i>	760	173	552.5	229.5
<i>Range</i>	239 - 1677	46 - 276	187 - 1292	34 - 746

Table 2. The granularity of the consonant inventory measurement in relation to the available speech data: group data of 30 Dutch-speaking two-year-olds.

<i>Granularity</i> (~incidence of consonants)	<i>high-frequent</i> ($p \geq 10\%$)	<i>mid-frequent</i> ($5\% < p \leq 10\%$)	<i>low-frequent</i> ($2\% < p \leq 5\%$)
<i>Size N of the required speech sample</i>	<i>29 tokens</i>	<i>59 tokens</i>	<i>149 tokens</i>
<i>Minimal size K of the entire speech session</i>	<i>at least</i> <i>32 tokens</i>	<i>at least</i> <i>61 tokens</i>	<i>at least</i> <i>151 tokens</i>
<i>Number of children's speech sessions containing ample consonant tokens</i>			
<i>Word position</i>	<i>initial</i>	<i>final</i>	<i>initial</i> <i>final</i> <i>initial</i> <i>final</i>
<i>Total</i>	30	30	30 26 30 20