



Universiteit  
Leiden  
The Netherlands

## **The development of the speech production mechanism in young children : evidence from the acquisition of onset clusters in Dutch**

Gulian, M.E.

### **Citation**

Gulian, M. E. (2017, October 31). *The development of the speech production mechanism in young children : evidence from the acquisition of onset clusters in Dutch*. LOT dissertation series. Retrieved from <https://hdl.handle.net/1887/57176>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/57176>

**Note:** To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/57176> holds various files of this Leiden University dissertation

**Author:** Gulian, Margarita

**Title:** The development of the speech production mechanism in young children : evidence from the acquisition of onset clusters in Dutch

**Date:** 2017-10-31

## Chapter 1. Introduction

### 1.1. Introduction

This thesis is about children's developing word production skills, and about the development of the system behind language production. The production of speech by adults has been studied in great detail, leading to several different models of the processes involved (Dell, 1986; Levelt, 1989; Levelt, Roelofs & Meyer, 1999; Boersma, 2011). However, up until now this line of research has hardly ever been extended to the (typically) developing speaker (cf. Wijnen, 1990; Stackhouse & Wells, 1997; Levelt, 1998). Despite the fact that child language productions typically deviate from the adult standard, the way the speech production mechanism performs and develops in the early stages of language production is largely unknown. In most work on phonological acquisition to date, some developmental state of the child's grammar is held responsible for these specific productions. However, the child language data that are studied are always production data; ignoring the real-time processes that have shaped these productions yields an incomplete account of the data (Docherty & Foulkes, 2000). We thus need to know more about the speech production mechanism of the developing speaker, and with this thesis I hope to contribute to this call.

I have limited the work in this thesis to a study of the system behind the production of isolated words, since this is what the developing speakers in this thesis, being between one and two-years old, mostly produce. Within the context of word-production, this study will focus on the - developing - production of word-onset consonant clusters. A typical deviation in early child language productions is the reduction of these clusters to singleton consonants, like in (Dutch) [tɛin] for target *trein* 'train', and [tup] for target *stoep* 'side-walk'. As mentioned above, up until now we only find grammatical accounts of this deviation, in the form of a fixed syllable template, a parameter setting, or a constraint on syllable structure (Fikkert, 1994; Pater & Barlow, 2003; Velleman

## 2 | Introduction

& Vihman, 2002). A brief discussion of these accounts will follow below in 1.4. However, instead of resulting from a specific grammatical setting, these cluster reductions could also be the outcome of the speech production process, and in the speech production mechanism there are several possible sources for error that could be considered. This is what will be done in this thesis, by studying children's cluster productions in different ways - acoustically, phonologically, and in relation to children's perception of consonant clusters - and analyzing both longitudinal, spontaneous production data, and elicited productions.

### **1.2 The speech production mechanism**

The different possible sources for error in child language productions that will be studied are the layers in the model depicted in Figure 1, based on the speech production model of Levelt, Roelofs and Meyer (1999) and the bidirectional model of Boersma and Hamann (2009), and Boersma (2011).

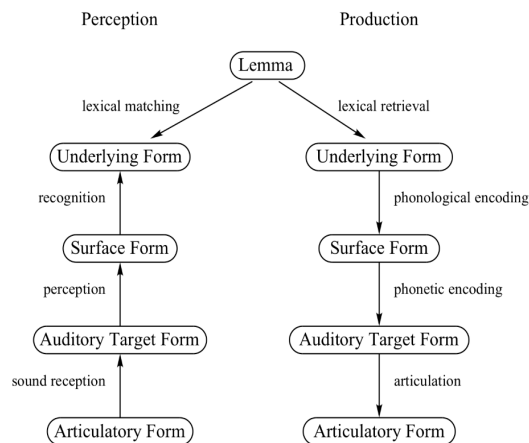


Figure 1. The perception-production model used in this study, elaborated on the basis of Boersma and Hamann (2009) and Levelt et al. (1999).

According to this serial processing model, and focusing first on the production side, in the mind of a speaker an intended concept is transformed in several steps into a motor program that will eventually be executed by the articulators. It takes around 600-700 ms from the moment of seeing a picture of a common object, like a train, to the moment of uttering the monosyllabic word *train* in a picture-naming task (Indefrey & Levelt, 2004; Szekely et al., 2004). In this very short time, the following steps have taken place:

1. Lemma activation (lemma = non-phonological part of an item's lexical information; Levelt, 1989). In the case of *train*, the lemma <train> will be activated.
2. Lexical retrieval. Each lemma activates its corresponding underlying, morphologically encoded, phonological form, which contains the stored information about the word's sounds, in this case /tren/, and the metrical frame, i.e. the number of syllables and stress pattern.

## 4 | Introduction

3. Phonological encoding. From this information, a phonological surface form is created. At this level, sounds are grouped into syllables, a single one in the case of *train*. I assume that this happens in a top-down way: segments are mapped onto stored syllable templates.

4. Phonetic encoding. Subsequently, the surface phonological form is converted into an auditory target form. In Levelt et al. (1999), it is assumed that for experienced speakers, motor programs for frequently-used syllables are stored in a mental syllabary, and can be retrieved directly. If a ready-made program (or the syllabary as a whole) is not available, the surface phonological form is provided with position-specific articulatory detail on the fly. In Levelt et al., the result of phonetic encoding is called the phonetic gestural score, but in Boersma and Hamann (2009) the phonetic encoding part is worked out in more detail, and is split into two modules, one that maps the surface phonological form onto an Auditory Target form, and one where this form is mapped onto an articulatory-motor program. Bite-block experiments have shown that speakers intend to produce vowels as closely as possible to an acoustic target, even when production is articulatorily inhibited (MacNeilage, 1981). This points to the existence of an auditory target form, which a speaker aims to achieve in production. The auditory target form is subsequently translated into an articulatory-motor program that controls the speech muscles. However, due to the limits of the present study, in this thesis, like in Levelt et al., a single phonetic encoding module is considered as possible error locus. Here, the phonological surface form is converted into the motor action instructions that will result in a form that the speaker aims to achieve in production, i.e. the auditory target form.

5. Articulation. The auditory target form is executed by the articulators, resulting in the acoustic realization of the word: [t.ɪ̃:n]

Although the main concern of this thesis is the speech production system, we need to take perception into account too. Speaking can hardly do without perceiving, decoding and representing speech. The model in Figure 1 includes this component. For word production, the focus of this study, the speech

comprehension system does not only play a crucial role in the way the sounds of words are stored - if certain sounds are not stored, they will certainly not be produced either - but also in what is called 'self-monitoring' by the speaker during the production process. Speech is monitored by the speaker before it is overtly articulated, as soon as a phonologically encoded form is available. For self-monitoring, the perception part of the model is used by the speaker, i.e. self-perception of inner speech takes place. If necessary, namely when an error is detected, repairs can be made before the speech is uttered. In the present study, I focus on perception only in relation to the segmental representations that form the input to the form-encoding part of word production. However, for a full understanding of the way developing speakers produce speech, perception and production and the systems underlying these processes should be studied in tandem. My hope is that as a sequel to the present work, the full model as depicted in Figure 1 above, will be studied in relation to phonological development.

### **1.3. Different sources of cluster reduction**

For the developing speaker, like for the mature speaker, all the different stages between lemma selection and actual articulation are potential locations for error, resulting in productions that deviate from the standard. For this study, it is assumed that the exact source of the error in the production mechanism can be deduced from the type of error that results. This, in turn, can inform us about the developmental state of (specific layers in) the mechanism.

If, for example, the target cluster is incompletely stored in the child's mental lexicon, with only one of the consonants, the error source is the underlying form, i.e. the segmental representation. In this case, we expect to find a highly systematic error; the consonant that is absent from the representation cannot be encoded in any way, so there will be a systematic and complete omission of this segment in the speaker's production. If, however, a target cluster is variably produced correctly and incorrectly, we can conclude that both

consonants of the target cluster are present in the segmental representation. An incorrect realization is then due to problems at lower levels of the production model, either at the level of phonological encoding or at the level of phonetic encoding. A single type of data is in general not enough to determine the exact error locus, and a combination of informative data needs to be considered. In Den Ouden (2002), an inspiration for the present study, the error locus in the production mechanism of patients with aphasia was determined on the basis of their performance on three different tasks, Picture Naming, Repetition, and Phoneme Detection. Arguing from the combined results of success on one task and failure on another, Den Ouden deduced whether the weakest link in the mechanism was formed by lexical access, phonological encoding or phonetic encoding. In Chapter 4 of this thesis, a similar procedure is used to find out about the development of the production mechanism.

#### **1.4. Phonological accounts of cluster reduction**

In phonological accounts of cluster development, usually two basic developmental stages are posited: an initial stage in which the underlying form /C<sub>1</sub>C<sub>2</sub>/ is reduced to a singleton [C] in the surface form – most commonly to C<sub>1</sub> if the target cluster consists of an obstruent followed by a sonorant – and a second stage in which a complete cluster can be present in the surface form, either correctly or with substituted segments. The initial stage, in which the cluster is reduced to a single C has been accounted for in different ways, and I will discuss the three most common ways here.

*Template account.* In this type of account, the child's production is constrained by a fixed template onto which consonants and vowels are mapped. Initially, this template is the core syllable, CV (Menn, 1976; Demuth & Fee, 1995; Demuth, 1996). An underlying representation /tren/ that is mapped onto this CV template, will end up as [te] in the surface form – and subsequently in production – because there are no positions available for the segments /r/ and /n/ in the template. This is shown in Figure 2.



Underlying representation:	/t r e n/
Template:	$\begin{array}{c}   \quad / \\ \text{C} \quad \text{V} \end{array}$
Output:	[te]

Figure 2: Cluster reduction in a Template account

*Parameter account:* Following the work by Chomsky (1981), Dresher and Kaye (1990) proposed a set of parameters governing the metrical structure of language. With respect to syllable structure, languages differ in their settings of parameters like the Minimal Onset Parameter ("Are Onsets obligatory?") and the Maximal Onset Parameter ("Can onsets be branching?"). In the initial stage of development, all parameters are in their default setting, and by paying attention to the input, the language learner will be able to change the default setting to the marked setting if evidence for this setting is present in the input. The default value for the Minimal Onset Parameter is *yes*, while for the Maximal Onset Parameter it is *no*. Together, these settings result in an initial grammar which only allows for syllables that have a single, obligatory consonant (Fikkert, 1994). In this initial stage, then, consonant clusters cannot be realized.

*Optimality Theory account:* In Optimality Theory (Prince & Smolensky, 1991), the phonological surface form results from an interaction of Markedness constraints, enforcing well-formedness, and Faithfulness constraints, enforcing the unaltered presence of information provided by the underlying form. The ranking of these constraints in a grammar determines the ultimate surface form of a specific underlying form. In the initial stage of development, Markedness constraints outrank Faithfulness constraints, and surface forms will thus have an unmarked, or well-formed, structure. Markedness constraints on syllable structure are Onset ("A syllable should have an onset"), No-Coda ("A syllable should not have a coda"), No-Complex-Onset ("A syllable should not have a

## 8 | Introduction

complex onset") and No-Complex-Coda ("A syllable should not have a complex coda"). Only CV syllables can be the output of the initial grammar, where all markedness constraints are ranked high (Gnanadesikan et al., 1995; Levelt, Schiller & Levelt, 2000).

In all three accounts, the phonological grammar enforces complete omission of one of the cluster consonants in the initial stage, and complete onset consonant cluster realization in the surface form in a subsequent stage, if required by the underlying form. Depending on the theory, development leading to the subsequent stage consists of the availability of a new template, CCV, the Maximal Onset Parameter setting changing from default *no* to *yes*, or a demotion in the ranking of the constraint No-Complex-Onset with respect to a Faithfulness constraint, allowing for violations of the markedness constraint. There are, thus, no intermediate forms of a target cluster in a grammatical account. In Chapters 2 and 3, however, we will encounter data that are difficult to explain in a grammatical account because the  $C_2$  is neither completely absent, nor completely present, or variably present or absent.

If we try to reconcile the phonological accounts with the psycholinguistic model, and a with a word production account, we could say that a grammar actually describes the limitations on the syllabification process in the phonological encoding module. This entails that if the problem with cluster realization lies in the phonological encoding module, we can expect complete, i.e. trace-less omissions of the underlying cluster segment  $C_2$  because there is no position for this consonant available in the syllable inventory that can be employed by phonological encoding. When we encounter data like in Chapter 2 and 3, where the target  $C_2$  is neither completely absent from, nor completely present in production, these are thought to result from flaws in the phonetic encoding module, or from a specific interaction between phonological and phonetic encoding.

### 1.5. Data

In this thesis, I have used both spontaneous and elicited data. In addition to studying production data, I carried out one perception experiment with young children (Chapter 5) and one with adults (Chapter 2).

The spontaneous word productions that I studied for Chapter 2 and Chapter 3 come from the CLPF database (Fikkert, 1994; Levelt, 1994) and are available through the CHILDES/Phonbank online database <http://phonbank.talkbank.org/> (Rose et al., 2006; Rose & MacWhinney, 2014). The CLPF corpus consists of spontaneous speech production data, of 12 children between 1 and 2 years of age at the start of a one-year data-collecting period, acquiring Dutch as their native language.

In addition, for the study in Chapter 2, I recorded 30 children with a mean age of 2;1 years at four Dutch day-care centers in the Amsterdam area, and for the study in Chapter 4, I used longitudinal data collected from four children who were between 1;7 and 2;1 years old at the start of the data collecting period in the Amsterdam area.

For the perception experiment in Chapter 2, thirty-five adult speakers of Dutch were tested, in order to find out whether they were able to discriminate reduced onset clusters from singleton onsets, produced by Dutch two-year-olds. For the perception experiment described in Chapter 5, fifty-eight children with a mean age of 2;0 were tested.

More specific information about the participants in every study is provided in the separate chapters.

### 1.6. Overview of the thesis

The study in **Chapter 2** concerns the question whether reduced clusters in children's productions are indeed fully reduced - warranting a phonological

## 10 | Introduction

account of cluster reduction - or whether they exhibit acoustic traces of the omitted second consonant. For this purpose, the acoustic characteristics of pairs of utterances, produced by the same speaker and at the same age, are compared, that differ only - or mostly - in the presence or absence of an onset cluster in their target forms, like *brood* /brɔ:t/ 'bread' - *boot* /bo:t/ 'boat' and *knip* /knɪp/ 'cut' - *kip* /kɪp/ 'chicken'. These words are realized in such a similar way that even trained phoneticians tend to transcribe them identically, e.g. as [bo:t] - [bo:t] and [kɪp] - [kɪp]. An acoustic analysis of these forms, however, reveals acoustic traces of the omitted consonants from the target clusters in the children's productions. The children in this study tended to realize a rising F2 in the vowel onset when the target C<sub>2</sub> was /r/, which might be reminiscent of the rising F3 that we see in adult speech. As for target words starting with /kn/, where /n/ was omitted from the production, we found that the subsequent vowel did show a moving formant pattern, and a lower center of gravity. In a subsequent perception experiment with adults, where they were presented with these semi-reduced utterances and their minimal pair counterparts, it turned out that these adult listeners could not decide which of the two productions referred to a target word starting with a consonant cluster.

In **Chapter 3**, we take a detailed look at the acquisition of clusters starting with a plosive and followed by /r/- hence /Cr/ - over time, by five different children, in their spontaneous speech. All their attempts to produce target /Cr/ clusters, from the start of the recording period until the cluster is produced correctly - or until the end of the recording period - are analyzed acoustically. Although the five children show individual developmental paths, a general pattern can be discerned; in Chapter 2 partially reduced clusters were found, here it is found that this type of realization forms a developmental stage, preceded by a stage in which complete omission of the C<sub>2</sub> takes place, and followed by stages in which the C<sub>2</sub> becomes more and more present and then becomes more and more correctly realized. The different stages are discussed in terms of developments in the speech production mechanism.

In **Chapter 4**, we look at the longitudinal performance of four children on three production tasks: PN (picture naming); WR (word repetition) and NWR (nonword repetition), where the target forms are real words or nonwords containing an onset cluster. Like in Den Ouden (2002), the functional state of the speech production mechanism is deduced from the combination of performance results on the different tasks. It is found that children perform poorly on the PN task in the initial sessions, while they do better on the NWR and/or the WR tasks. This points to the lexical representation as the initial error locus because performing successfully on the NWR and/or the WR task does not require lexical access. In later sessions, the error pattern changes. Like in Chapter 3, these changing error patterns are taken to reveal developments in the speech production mechanism, and they are discussed in detail.

In **Chapter 5**, I turn to perception, and ask how detailed the representation of onset clusters is in the child's mental lexicon. Do children exhibit different looking behavior when they perceive correctly produced onset clusters as opposed to reduced onset clusters? If this is the case, the segmental representation can be assumed to be detailed, containing both  $C_1$  and  $C_2$ . If not, omissions in production could be the result of incomplete segmental representations. I examine two-year-olds' perception of correct vs. reduced /sC/ clusters, like in the word *stoel* /stul/ 'chair' and /C+liq/ clusters like in the words *trein* /trɛin/ 'train' and *bloem* /blum/ 'flower'. Interpreting the looking times in line with earlier work on children's perception of mispronounced words (Swingley & Aslin, 2000, White & Morgan, 2008), results seem to indicate that two-year-olds exhibit awareness of /sC/ cluster reduction but not of /C+liq/ cluster reduction. However, another interpretation of the results is that the longer looking times actually indicate that the correct form is novel to the child, and therefore attracts longer attention. This interpretation is strengthened by the children's performance on a small production task, where they simply had to name the pictures that were shown in the perception

## 12 | Introduction

experiment, and where we find that in children who have not acquired /sC/ clusters yet, this novelty effect is stronger than in children who have already acquired /sC/ clusters.

Finally, in **Chapter 6**, the results obtained in Chapters 2 to 5 are discussed in relation to each other, and I will summarize what the combination of results can tell us about the developing speech production mechanism.



