UTRECHT UNIVERSITY

Faculty of Humanities

Research Master Linguistics

Computational Modeling of Error Patterns in Children Speech

First Supervisors:
Prof. Dr. Wijnen & Dr. Kroon

Candidate:
Alex Stasica

Second Reader:
Dr. Nazarov

In cooperation with:
Auris

July 15, 2024

Abstract

This study aims to analyze the phonological processes in the speech of Dutch children, focusing on both typically developing (TD) children and those with developmental language disorder (DLD). Utilizing a dataset from a non-word repetition task, we investigate some phonological errors made by children aged 3;0 to 6;2 (years; months). Our approach involves using a combination of Levenshtein Distance and Breadth-First Search algorithms to quantify and document four common phonological processes (ie. error patterns): final consonant deletion, stopping, fronting, and gliding.

We perform statistical analyses to compare the frequency of these processes between TD and DLD children and to assess differences in the percentage of pseudowords presented and repeated by each group. Building on this analysis, we apply Optimality Theory constraints and train a maximum entropy (MaxEnt) model to evaluate each child's pronunciation. This model is trained on TD children's data and tested on both TD and DLD children to determine the probability of typical pronunciation patterns.

The effectiveness of the classifier is assessed using receiver operating characteristic curves to distinguish between TD and DLD children. Our findings indicate significant differences in the phonological processes use and number of words repeated (either correctly or incorrectly) between the two groups, supporting the utility of these metrics in diagnosing DLD. Additionally, the MaxEnt model demonstrates high reliability especially in the oldest group of children.

This research contributes to clinical practice by offering detailed analyses of child pronunciation errors and improving diagnostic accuracy for DLD. Theoretically, it advances our understanding of phonological development and the applicability of OT in language acquisition studies.

Contents

1 Introduction **4**

2 Methods Overview **7**

3 State of the art **10**

3.1 Phonological Acquisition in Monolingual Children 10

3.2 Dutch Phonology 14

3.3 Language Disorders and Phonological Acquisition 15

3.4 Computational Modeling Approaches of Phonological Acquisition and Diagnosis 23

4 Research Questions **37**

5 Methods **38**

5.1 Data 38

5.2 Phonetic Transcription 39

5.3 Phonological Processes 42

5.4 Automatic Detection of Phonological Processes 43

6 Results **64**

6.1 LD-BFS output results 64

6.2 MaxEnt results 69

7 Discussion **74**

7.1 Limitations and further work 74

7.2 Conclusion 77

A Appendix A **80**

B Appendix B **82**

C Appendix C **84**

Bibliography **93**

1. Introduction

This master's thesis aims to be a first step to the resolution of a gap in computational analysis of phonological development in children across different age groups. Specifically, the focus is on discerning differences between typically developing (TD) children and those with developmental language disorder (DLD) through the analysis of four key error patterns.

One key aspect of studying language development in children, both TD and those with DLD, is the analysis of phonological processes (i.e., phonological error patterns resulting from simplifying the speech of children during their acquisition). Analyzing these errors provides valuable insights into the underlying mechanisms of language acquisition and potential developmental issues. However, manual assessments are particularly time-consuming and current automatic techniques for detecting and categorizing them often lack precision and specificity. This makes it challenging for speech and language pathologists (SLPs) to accurately diagnose and treat children with language disorders.

Improving the automatic detection and categorization of phonological processes (PP) based on age and impairment status is therefore crucial for enhancing the efficacy of diagnostic processes and intervention strategies for children with DLD. Early identification and targeted intervention can significantly improve long-term outcomes for these children. By addressing these challenges early on, they can better develop their personal skills, build stronger social connections, and strengthen their self-esteem.

Currently, the Klank Analyse Tool (KAT), developed by Auris—an organization dedicated to assisting individuals with speech and language difficulties—is used for automatically analyzing Dutch children's speech and PPs. However, KAT relies on 'pseudo-phonetic' transcription using graphemes provided by a human annotator, which are likely to introduce

biases into the analysis.

The tool operates as follows: the SLP prepares a set of reference words for a specific task, such as non-word repetition (NWRT) or picture naming task. The SLP records the child's performance and then fills in the target word and the word as pronounced by the child in pseudo-phonetic script.

In this process, the SLP phonetically transcribes what the child pronounced using an orthographic keyboard, with guidelines to dictate how to fill the 'pronunciation' column. KAT then generates different spreadsheets with various types of phonological analyses based on this information.

However, as stated above, a significant limitation of this method lies in the use of orthography for transcription, which can introduce a lack of precision. For instance, the influence of orthographic rules may lead to inaccuracies or representing the same phonemes inconsistently (i.e., with different orthographies). Additionally, the time required for SLPs to manually fill in information for numerous children and words poses a practical challenge.

Therefore, the objective of this research project is to develop a pilot pipeline aimed at overcoming KAT's limitations. By using computational techniques and machine learning algorithms, the project seeks to develop more accurate and efficient methods for identifying and analyzing these pronunciation errors. In addition, the project aims to provide support for diagnosis by classifying the child as typical or with DLD. This classification can serve as a valuable tool for SLPs in their decision-making process, aiding them in diagnosing and treating this disorder more effectively. These advancements aim to facilitate the work of researchers who study child phonology.

In this research, we study these challenges in children aged 3;0 to 6;2 (years; months) because of the nature of errors in early language production (i.e., younger children's errors cannot be reliably categorized [1]).

The structure of this thesis is as follows: Section 2 presents an overview of the methods used during this research along with the framework in which the models are developed and their relevance for clinical practice and

for the field of language acquisition. Section 3 presents an overview of the current state of research in language acquisition. It focuses on phonological acquisition for all children acquiring their first language, with special reference to the acquisition of Dutch. It also reviews current computational models designed to better understanding phonological acquisition and supporting SLPs. Section 4 outlines the aim of the present study and the research questions. Section 5 details the methodology. Section 6 presents the results. Finally, section 7 discusses the limitations of this research and suggests avenues for future work.

2. Methods Overview

Code and Data Availability

In this research, we use a dataset from a Non-Word Repetition Task (NWRT) performed by typically developing (TD) children and children with developmental language disorder (DLD) between ages 3;0 and 6;2 (years;months) [2]. In the dataset used, each target pseudo-word is paired with its transcribed pronunciation by the children.

The program developed for this study is accessible on Google Colab at `https://colab.research.google.com/drive/15xndGrSwghQ31lMyF1oJNVkOk6Vp3_dW?usp=sharing`. The dataset utilized in this study is publicly available, as all children's data have been anonymized, and no audio recordings were used. The dataset does not contain any sensitive information. The Excel file with the anonymized data is hosted on Google Drive at `https://drive.google.com/drive/folders/121uOHPOTUggvkIudvwUrSkwtfkxRc8Zz?usp=sharing`.

Theoretical Framework

All pronunciation errors are analyzed as part of the developing phonological grammar of the child, and not as part of other components involved in pronunciation and presenting a late development, following the perspective of the Optimality Theory (OT) [3].

Phonological Processes Analysis

Using the NWR data with target and pronounced words, we initially apply a combination of the Levenshtein Distance and Breadth-First Search algorithm to identify and quantify the phonological processes employed by the children. The results of this initial analysis provide a detailed summary

of the phonological processes used by each child. For this pilot research, we focus on four common phonological processes: final consonant deletion, stopping, fronting and gliding. We document in which words each phonological process occurs, which phonemes are affected, and summarize the percentage of use of each phonological process by child.

Furthermore, we perform statistical analyses to identify if there is a significant difference in the frequency of use of the four phonological processes studied here between TD children and those with DLD present in our dataset. By understanding how these phonological processes vary between TD children and those with DLD, we can better interpret the patterns of pronunciation errors. This can help in refining the classification models used for diagnosing DLD in future research, ensuring that the models accurately reflect the differences in phonological development between the two groups.

We also perform a statistical analysis to determine if there is a significant difference in the number of words repeated (either correctly or incorrectly) by TD and DLD children. This analysis is crucial because the number of words pronounced can be an indicator of language proficiency and development. A significant difference in this metric between TD and DLD children would support the validity of using the percentage of pronounced words as a diagnostic feature.

Child Support Diagnosis

Building upon this detailed phonological analysis, we translate these phonological processes into OT constraints. Each target form - child's pronunciation pair is evaluated for its adherence or violation of these constraints using a maximum entropy (MaxEnt) model. We train separate models for each age range (3-4, 4-5, and 5-6 years old) using a training set consisting only of TD children. Subsequently, we test the models using a test set containing both TD children and children with DLD. The MaxEnt model outputs a probability for each pronounced word, representing the likelihood of it matching typical pronunciation patterns observed during

training.

To categorize each child as TD or with DLD, we calculate the mean probability across all repeated words for that child. A higher probability indicates greater similarity to typical pronunciation patterns seen during training, aiding in determining if the child exhibits typical development or signs of DLD. We use receiver operating characteristic (ROC) curves to determine the optimal threshold probability for distinguishing between typical and atypical children.

Theoretical Relevance

This research aims to be relevant to both clinical practice and theoretical linguistics. The method enables the detailed analysis of pronunciation errors in children, facilitating a precise assessment of language development and potential disorder. This information is expected to support speech and language pathologists in making informed clinical decisions regarding diagnosis and intervention strategies.

From a theoretical perspective, this research contributes to the field of linguistics by empirically testing the application of OT in the context of language acquisition. OT provides a framework for understanding how phonological systems develop and organize across languages. By applying OT constraints to analyze children pronunciation patterns, this study explores universal principles underlying phonological development.

The approach adopted here utilizes straightforward and resource-efficient methods that prioritize interpretability, making it accessible for both researchers and practitioners in linguistics and clinical settings. This methodological clarity enhances the reliability of analyzing child error patterns, thereby advancing theoretical insights into language acquisition processes.

3. State of the art

3.1 Phonological Acquisition in Monolingual Children

Phonological acquisition is a staged process. It begins with the initial acquisition of the segmental inventory, which includes learning the individual sounds (phonemes) of a language. This is followed by the acquisition of segmental rules or processes, which are the systematic patterns or transformations that govern how these sounds are produced and altered in different contexts. Ultimately, phonological acquisition involves integrating both the segmental inventory and these segmental rules or processes [4].

During the segmental processes acquisition stage, children begin to adapt to adult speech forms, moving beyond their initial production constraints. This phase has been described as the start of systematic modifications in reproducing adult speech segments, sequences, and syllable or word structures [5]. Figure 3.1 illustrates these systematic changes through the stages of a Dutch child's acquisition of plosive-liquid clusters.

Ingram was among the first to characterize typical child phonological rules across languages, identifying the systematic nature of child phonological behavior [6]. He emphasized the importance of examining four distinct forms: adult pronounced word, child pronounced word, child perceived word form, and child underlying form. Understanding these forms is crucial for a comprehensive analysis of child phonological processes (PPs). While the adult and child pronounced words can be directly observed through transcription and recording, the child's perceived word form and underlying form require careful inference from patterns observed in the child's speech production and systematic errors.

a.	STAGE 1: *Cluster reduced to plosive*			
	broek 'trousers' /bruːk/	→	[buːk]	(1;9.21)
	trein 'train' /trɛin/	→	[kɛŋ]	(1;9.21)
	klaar 'ready' /klaːr/	→	[kaː]	(1;10.15)
	plassen 'to pee' /ˈplɑsə(n)/	→	[ˈpaːsə]	(1;10.15)
b.	STAGE 2: *Cluster reduced to sonorant*			
	drinken 'to drink' /ˈdrɪŋkə/	→	[ˈliːkɛ]	(1;10.29)
	klap 'bang' /klɑp/	→	[lɑp]	(1;10.29)
	klok 'clock' /klɔk/	→	[lɔk]	(1;11.12)
c.	STAGE 3: *Cluster produced as cluster*			
	bloemen 'flowers' /ˈbluːmə/	→	[ˈbluːmɛ]	(1;11.12)
	kleur 'color' /kløːr/	→	[klœ]	(1;11.12)

Figure 3.1: Plosive-Liquid cluster [4]

Several works in language acquisition presuppose that differences between child and adult forms are caused by the phonological system. Two contrasting perspectives exist within this framework. Phonological processes can be viewed as realization rules that are later unlearned [7], or as constraints on output production, with development involving the removal of constraints and/or elaboration of templates to align the child's form more closely with the adult target [1]. The present study adopts the latter approach, specifically using the framework of Optimality Theory (OT)[3]. OT distances itself from rule-based phonological systems, focusing instead on constraints on output that are progressively eliminated in typically developing (TD) children but may persist longer in children with developmental language disorder (DLD) or delayed development.

To further understand these systematic differences between child and adult forms, Ingram categorizes the systematic errors observed in children's speech. He identified three main types: syllabic processes, assimilatory processes, and segmental substitution processes [8]. This framework has been widely used to explore various PPs in child language (e.g.[9] [1] [10]). Key PPs, and their definitions include:

- **Syllabic processes: Involving complexity reduction**
  - Final consonant deletion: omission of the final consonant of a word
  - (Unstressed) syllable deletion: elimination of syllables, particularly preceding a strong syllable (i.e., one with a longer vowel, louder volume, and higher pitch) or if unstressed
  - Cluster reduction: simplification of consonant clusters by removing one or more consonants, resulting in a single consonant in the pronounced word
  - Reduplication: repetition of one or more syllables within a word
- **Assimilatory processes: Changes in phonemes due to the features of neighboring segments**
  - Harmony: influence of one sound on another within a word, encompassing velar, nasal, labial and voicing assimilation
- **Segmental substitution processes: Involving replacement of sound segments by others**
  - Fronting: replacement of a consonant with one articulated further forward in the oral cavity
  - Stopping: substitution of fricatives with stop consonants at corresponding places of articulation
  - Gliding and vocalization: substitution of liquids with glides or vowels

The above list outlines the most frequently observed PPs, yet this list is non-exhaustive and other processes are used by the children based on the language they acquire and their age [7].

Adding to Ingram's categorization [6], other researchers worked on providing a comprehensive overview of these PP and the age ranges in which they are used [11]. Such an overview is summarized in Figure 3.2, which outlines the principal PPs and the typical sequence in which they emerge

during development. The table illustrates different age ranges, the typical progress in segmental acquisition, and the PPs found at each age range. PPs are in upper-case to indicate their presence at a given age, with optionally present or infrequent PPs in parentheses or in lower-case. While PPs may occur outside this sequence, persistent occurrence beyond the typical age range(s) may indicate atypical development. Generally, processes affecting the syllabic structure are prevalent until around age three in TD children, diminishing thereafter, while processes affecting individual segments persist longer in the acquisition process.

Stage I (0;9–1;6)	Nasal Plosive Fricative Approximant	Labial		Lingual	'*First Words*' tend to show: —individual variation in consonants used; —phonetic variability in pronunciations; —all simplifying processes applicable.	
Stage II (1;6–2;0)	m p b w	n t d			Reduplication Consonant Harmony FINAL CONSONANT DELETION CLUSTER REDUCTION	FRONTING of velars STOPPING GLIDING /r/→[w] CONTEXT SENSITIVE VOICING
Stage III (2;0–2;6)	m p b w	n t d		(ŋ) (k g) h	Final Consonant Deletion CLUSTER REDUCTION	(FRONTING of velars) STOPPING GLIDING /r/→[w] CONTEXT SENSITIVE VOICING
Stage IV (2;6–3;0)	m p b	n t d		ŋ k g	Final Consonant Deletion CLUSTER REDUCTION	STOPPING /v ð z ʧ ʤ/ /θ/→[f] FRONTING /ʃ/→/[s] GLIDING /r/→[w] Context Sensitive Voicing
Stage V (3;0–3;6)	f w	s (l)	j	h	Clusters appear: obs.+approx. used /s/ clusters may occur	STOPPING /vð/ (/z/) /θ/→[f] FRONTING of /ʧ ʤ ʃ/ GLIDING /r/→[w]
Stage VI (4;0–4;6) (3;6–4;0)	m p b f v w	n t d s z l (r)	ʧ ʤ ʃ j	ŋ k g h	Clusters established: obs.+approx. 'immature' /s/ clusters: /s/→Fricative obs.+approx. acceptable /s/ clusters: '[s] type' fricative	/θ/→[f] /ð/→[d] or [v] (PALATALIZATION of /ʧ ʤ ʃ/) GLIDING /r/→[w]
Stage VII (4;6<)	m p b f v θ ð w	n t d s z l r	ʧ ʤ ʃ ʒ j	ŋ k g h	(/θ/→[f]) (/ð/→[d] or [v]) (/r/→[w] or [ʊ])	

Figure 3.2: Profile of phonological development for English [11]

While the development patterns discussed in this section are generally applicable across different languages, variations exist due to the unique phonological characteristics of each language. These variations influence the frequency and application of PPs across languages. In the context of

Dutch language acquisition, specific PPs have been identified by researchers (e.g., [12] [13]).

3.2 Dutch Phonology

While the preceding section provided a broad overview of phonological acquisition and processes in children, this section narrows the focus to Dutch phonology.

3.2.1 Dutch Adult Phonology

Dutch phonology has been studied extensively in research [14]. It has been noted that Dutch comprises 23 consonants, with additional ones appearing in loan words, and 17 vowels, including 3 diphthongs.

Dutch phonology also exhibits specific phonotactic constraints. Syllabic structures can consist of zero to three consonants in the onset and up to four consonants in the coda, with vowel-only syllables also permissible. The acquisition of these different syllable structures has also been described [15].

Regarding consonants and consonant clusters, specific restrictions apply, varying depending on whether they occur in syllable-initial (SI) or syllable-final (SF) positions.

- The phoneme /ŋ/ cannot occur at the SI position
- The phoneme /h/ is prohibited in syllable-final position and cannot be part of a cluster
- SI clusters cannot consist of two sonorant consonants, except in words of foreign origin
- SI clusters with three consonants must have /s/ as the initial consonant

- SF clusters with more than two consonants must contain both /s/ and /t/

3.2.2 Dutch Phonological Development

During the phonological development of Dutch speaking children, vowels, including diphthongs, are typically mastered by age 3;0. By age 4;0, most single consonants are correctly produced in initial and final positions by 75% of children, with the exception of /s/ and /r/, and by age 4;3, most sounds, including single consonants and consonant clusters, are articulated correctly by the majority of children.

Furthermore, as already detailed above, children exhibit phonological processes during their acquisition, and previous research has determined at which age these processes can be expected to stop. Section 5.3 details the specific PPs examined in this research, including the age at which they typically cease to occur. The age ranges of use for each PP serve as benchmarks in assessing children's developmental typicality during evaluations of their speech production at different stages. It is recognized that occasional incorrect productions may occur within these age-appropriate ranges [1].

3.3 Language Disorders and Phonological Acquisition

Children with developmental language disorder (DLD) may not always exhibit phonological issues[1], or they might experience these issues at varying levels of severity. Understanding the frequency and consistency of these phonological issues is crucial for assessing a child's phonological development.

In this context, Vihman categorizes the relative frequency of phonological processes (PPs) as sporadic (<25%), inconsistent (25-75%), or regular

[1] In contrast to DLD, another clinical entity called 'speech sound disorder' is constently associated with phonological issues during a child's development.

(>75%) [5]. This categorization remains relevant today and is useful for the assessment of a child's phonological development, providing a structured way to measure and compare the occurrence of phonological processes.

Building on Vihman's framework for categorizing phonological processes, numerous studies have investigated the different word production errors and their frequency in children with DLD in diverse languages and also compared them to typically developing (TD) children. Marshall, for instance, summarizes various findings regarding the production differences between children with DLD and TD children [16]. Notably, children with DLD make systematic errors in existing words during production, and repeating nonsense forms. Models of word production must account for these systematic errors in a cohesive manner.

Focusing on the use and frequency of PPs by Dutch children, Beer presents a study involving the recordings of spontaneous speech samples from 15 Dutch children with DLD from 4;0 to 6;0 (years; months), comparing them to TD children and to Swedish and English children with DLD [17]. She divides the PPs in three categories, a) the ones found in normal phonological development, which have a negative effect on intelligibility if they persist, b) unusual processes, appearing in normal development but infrequently and c) the ones virtually absent in children with typical phonological development, also referred as 'idiosyncratic' by [11]. Within each category, she differentiates the phonotactic processes (i.e., affecting the sequential structure) and the ones simplifying the system of contrasts (i.e., affecting the segments). The PPs studied in her research [17] are duplicated in Table A.1 in Appendix A.

Beers' findings [17] indicate that there are no specific PPs unique to children with DLD. Instead, children with DLD tend to exhibit higher frequencies of common PPs. Furthermore, cross-linguistic comparisons reveal that children with DLD use PPs with similar frequencies across languages, except for a few that are language-specific.

Though Beers uses spontaneous speech sample in her study, it is common to use non word repetition tasks (NWRT) to analyze the PPs of children

with DLD. Indeed, the repetition of pseudowords is particularly challenging for them. Analyzing how these children repeat pseudowords, and the types of errors they make, can reveal crucial information about the phonological aspects of their word production system. This task captures perception, storage, and reproduction of phonological forms. Using an NWRT for diagnosing DLD is advantageous because it requires only a few target words with specific sound combinations to assess a child's phonological abilities. In contrast, analyzing spontaneous speech necessitates more extensive samples to represent the diverse types of errors made by the child. This means that NWRT can efficiently pinpoint the phonological processes and errors, providing a clear diagnostic insight into the child's phonological memory and production system. Thus, NWRT complements the findings from spontaneous speech by offering a targeted approach to identifying phonological deficiencies in children with DLD.

Different phonological theories aim to explain the phonological deficiency of children with speech disorders, including DLD. The theory followed by the present study is the Optimality Theory (OT) [3]. In order to understand the most widely used framework of phonological impairment within OT, it is important first to present how OT works.

3.3.1 Optimality Theory Principles

OT emerged as a response to the limitations of rule-based systems, particularly evident in the framework of generative phonology outlined in the Standard Theory of Generative Phonology [18]. While generative phonology offers significant insights, it struggles to provide a comprehensive explanation of how phonological systems of the different languages can be explained by the same processes, as generative phonology describes different rules for the phonology of each language [19].

This gap in explanation spurred the development of OT, which holds promise not only in understanding phonological systems but also in explicating phonological development and disorders [20].

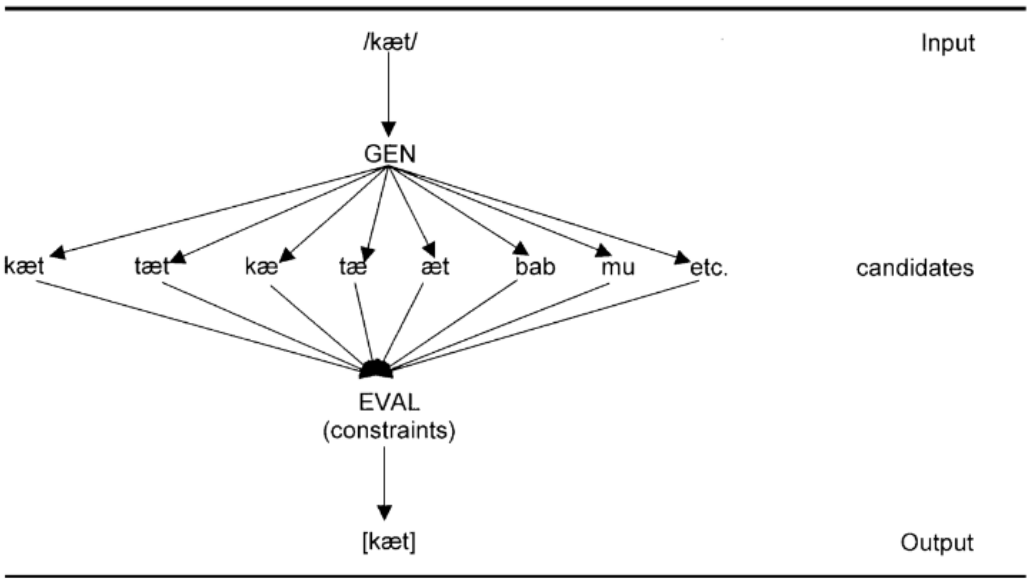

Figure 3.3: Schematic of Optimality Theory [20]

One of the core principles of OT is its treatment of the disparity between mental representation (i.e., underlying phonemic structure) and surface representation (i.e., spoken phonetic form). Previous theories attributed this difference to distinct rules, with variations between languages attributed to different rule sets. In contrast, OT posits that these differences arise from the varying rankings of constraints.

Formally, the OT model consists of three main components, as described below [21], accompanied by a schematic representation in Figure 3.3:

- **GEN (Generator)**: This component generates an infinite set of potential output (i.e., production) forms based on a given input (i.e., mental representation).
- **EVAL (Evaluator)**: Given the candidate set produced by GEN, EVAL selects the optimal output form considering the input representation.
- **CON (Constraints)**: EVAL utilizes a language-specific ranking of universal constraints to determine the optimal output form.

OT operates with two major families of constraints:

- **Markedness Constraints**: These constraints pertain to the well-formedness of the output and aim to reduce structural complexity and contrast between words by imposing restrictions on the output.
- **Faithfulness Constraints**: These constraints focus on preserving structural elements and aim to prevent deviations between the input and output.

The inherent conflict between these constraint families leads to constraint violability. Each output violates certain constraints, and the chosen output—considered the optimal—minimizes violations of the highest-ranked constraints compared to competing candidates.

Central to OT is the notion of optimality, where the selection process by EVAL aims to choose the optimal candidate. In this context, no output is inherently good or bad; rather, the optimal output is the one that minimizes violations of the highest-ranked constraints.

Notably, not all constraints are active for a given input. A constraint is considered active if it plays a decisive role in distinguishing between potential output candidates. Specifically, a constraint becomes active when it discriminates among several output possibilities that have not already been ruled out by higher-ranked constraints. Inactive constraints either do not apply to a given input and its potential outputs, or they apply but do not influence the selection process because higher-ranked constraints have already eliminated all potential candidates they could affect.

While OT typically aims for a single winner-defined as the output with minimal constraint violations—, cases of multiple winners can occur, indicating equal ranking of conflicting constraints. This aspect of OT accommodates output variability, such as in the phonological development of children. For instance, when children learn to produce sounds, they may sometimes produce multiple acceptable variants of a word due to their developing phonological systems. These variations can be understood within

the OT framework as instances where different constraints are equally balanced, resulting in more than one optimal output. This accounts for why children's speech may show variability as they navigate and resolve these conflicting constraints during their language acquisition process.

3.3.1.1 Exploring the Application of Optimality Theory in Phonological Development and Disorders

Understanding phonological acquisition requires consideration of various factors [22]. These factors include the discrepancy between a child's production and adult input forms, the variability observed both within and across developing systems, and the development of a child's grammar over time. In the framework of OT, such discrepancies signify different constraint rankings not only between adults and children but also among individual children, showcasing extensive variation.

OT posits that language acquisition begins with an initial structured state, guided by the principles of Universal Grammar (UG). This initial state refers to a universal set of constraints that are present from the beginning in all children. These constraints are initially ranked in a particular order, which may not yet resemble the adult target grammar. Children navigate through linguistic input aided by UG, gradually constructing a grammar resembling that of the adults around them. Acquisition occurs primarily through positive evidence present in the learner's input.

Children's early linguistic output tends to exhibit simpler forms compared to adults [23] [24]. This simplicity reflects the prevalence of unmarked structures in the initial stages of language acquisition. OT provides an explanatory framework for this phenomenon by positing that markedness constraints, which favor simpler and more universal structures, initially dominate over faithfulness constraints, which require maintaining the specific details of the input. As a result, children first acquire the less complex, unmarked structures. Over time, as they receive more linguistic input, the ranking of constraints adjusts, allowing for the acquisition of more marked

and complex structures. This dynamic interaction between constraints explains the observed progression from simpler to more complex forms in children's language development.

In adult language, a balance between marked and unmarked structures is essential to support a diverse lexicon. Even though markedness constraints may be dominated in adult grammar, they remain active when they do not conflict with dominating faithfulness constraints.

The ranking of constraints in a child's developing grammar dictates the error patterns exhibited. If an output displays multiple error patterns, it signifies the influence of high-ranking markedness constraints that require deviations from the input in various aspects. There are two types of variation; intra-word and inter-word variation [20].

Intra-word variation occurs when the surface form of a word varies within a single grammar (i.e., in one child). This means that for a given word, there may be multiple possible pronunciations or phonological realizations. In OT, intra-word variation is typically attributed to constraints that are ranked equally within the same stratum. When constraints are equally ranked, they can compete to determine the optimal output for a specific word. Therefore, different instances of the same word may exhibit slight variations in pronunciation due to these competing constraints. For example, a child pronouncing several times the same word can sometimes exhibit a phonological process and sometimes not.

Inter-word variation refers to differences in how phonological rules or constraints are applied across different words or contexts in a child's speech. Unlike intra-word variation, which focuses on variability within a single word, inter-word variation involves variations observed between different words in a child's production.

In OT, inter-word variation typically arises from constraints that are ranked across different strata. This means that constraints from different levels of hierarchy (strata) interact to influence how phonological rules are applied across different lexical items or syntactic contexts. For instance, a child may produce some words with a phonological process and other

words without the same phonological process, depending on how constraints at different levels prioritize the preservation or the violation of the input structure.

3.3.1.2 Using OT analysis to determine treatment goals

The most widely applied implementation of OT in phonological acquisition considers that reranking constraints play a crucial role in determining treatment goals, marking a departure from previous theories that attributed development solely to suppressed phonological processes or lost rules. This OT view considers that children's mispronunciations are due to them having a different phonological system than adults, and not to performance limitation.

In this context, 'performance' refers specifically to the various systems involved in producing or interpreting language (e.g., physical mechanisms of oral production). The OT framework assumes that developmental errors in phonological acquisition are only due to differences in the underlying phonological representations and constraints within the child's linguistic system, with different ranking of constraints between adults and children in their phonological grammar. These errors are not considered errors of language use in the broader sense that encompasses everyday variability in speech production observed in both children and adults.

Understanding what triggers constraint reranking remains a question within OT, although positive evidence likely plays a significant role. An effective framework must accommodate changes in grammar over time, as observed in TD, where grammar evolves to allow for the demotion of markedness constraints [25] [20] [23] [24].

For children with DLD, SLPs must provide explicit positive evidence, much more densely than is normally done by the caretakers, to facilitate grammar development. The clinician's objective is to induce the demotion of high-ranking markedness constraints. Identifying these constraints enables the targeting of specific linguistic features for intervention. Stud-

ies have demonstrated the effectiveness of introducing marked structures into a child's phonological system, resulting in comprehensive system-wide changes [26] [27].

3.4 Computational Modeling Approaches of Phonological Acquisition and Diagnosis

While the preceding sections explored phonological acquisition in children, this section examines various computational models that simulate phonological development and assist in diagnosing different speech disorders in children. It is important to clarify that this subsection does not exclusively focus on developmental language disorder (DLD). Instead, it reviews approaches used to model typical phonological development and its deviations. Although the presented models may not specifically address DLD, they offer valuable insights into effective modeling techniques for understanding children's phonological development.

3.4.1 Phonological Processes and Phonotactic Learning Modeling

3.4.1.1 EPAM-VOC: Model for Phonotactic Learning in English

Given that children with DLD typically perform poorly on non-word repetition tasks (NWRT) [16], the EPAM-VOC (Elementary Perceiver and Memorizer - Vocabulary) model has been proposed [28]. It has been adapted to simulate error patterns in TD children's word learning processes based on NWRT performance, particularly to replicate the higher frequency of errors in syllable onsets compared to syllable codas [29].

EPAM-VOC organizes a child's knowledge of sounds into a structured hierarchy, similar to a family tree, with individual sounds at the top and combinations of sounds forming words at the lower levels. This structure is illustrated in Figure 3.4.

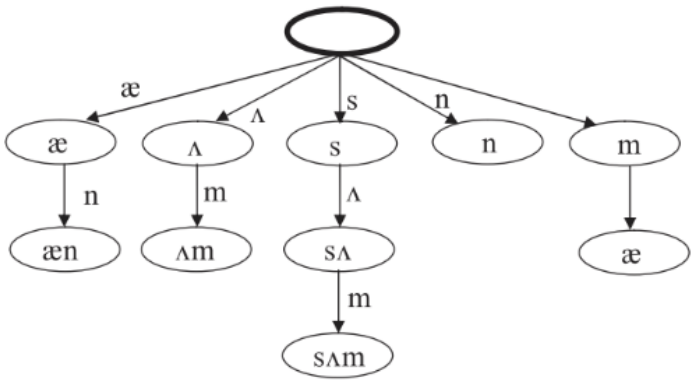

Figure 3.4: Illustration of the EPAM-VOC architecture. Nodes are represented by ellipses and links by arrows [29]

The model builds a network to recognize and categorize sounds by analyzing features of the sounds it hears, and then storing this information as long-term knowledge.

This model has been adapted for English by including all the sounds used in the English language right from the start, helping the model focus on the relevant features [29]. This is why they focused on children older than 5;0, as children of this age are expected to know all the sounds of their native language.

EPAM-VOC also investigates how children learn new words and store their sound patterns in their short-term memory, which can hold information for about 2000ms. The model examines how this short-term memory interacts with long-term knowledge to understand why some children learn words faster than others. The limited capacity of short-term memory acts as a bottleneck, controlling how much information can be stored in long-term memory. This interaction helps explain differences in how quickly children can learn new words. The performance of children on NWRT is used to test this model, allowing exploration of how effectively children can store and process unfamiliar sound patterns.

3.4.1.2 Klank Analyse Tool: Phonological Processes Analysis for Dutch Speaking Children

As mentioned in the introduction, KAT, a model for phonological processes (PP) identification in Dutch children has been developed by Auris. The computational details of the model are not publicly available, but the tool aims to support the speech and language pathologists (SLPs) when analyzing the performance of a child on a specific task. The tool operates as follows: the SLP prepares a set of reference words or pseudowords for a specific task, such as NWRT or picture naming. The SLP records the child's performance and then fills in the graphemic transcription of the target word, its pseudo-phonetic transcription and the pseudo-phonetic transcription of the word as pronounced by the child.

KAT also provides guidelines to dictate how to fill the 'pronunciation' column, with mandatory rules (e.g., how to divide each word in syllables, and how to transcribe syllable deletions or insertions) and optional rules (e.g., different possibilities on how to transcribe the schwa or on how to show distortions and striking sounds in the child's pronounciation).

Then, KAT generates multiple spreadsheets containing various types of phonological analyses derived from this information. These include spreadsheets detailing the percentage of correct pronunciations per sound, identifying instances where sounds have been correctly produced and their frequency. Additionally, for incorrectly pronounced sounds, the spreadsheets provide explanations of the errors. Other spreadsheets cover the phonological processes observed, list all words containing clusters, and calculate the percentage of correct cluster pronunciations. These spreadsheets also offer the capability to compare results across different tests.

As outlined in the introduction section, this tool has limitations and the present study aims to overcome these limitations and create a model capable of identifying PPs and their frequency given a real phonetic transcription.

3.4.2 Modeling Approaches for Speech Disorder Diagnosis

3.4.2.1 Computer-Aided Speech Therapy (CAST) Tools

Numerous CAST tools have been been developed as practical alternatives to human assessment, assisting in diagnosis and treatment planning for individual children [30]. Early detection of the various speech disorders children can have is crucial as it directly impacts fluency and intelligibility, emphasizing the importance of timely intervention.

The main types of pronunciation errors observed in children with speech disorders have been categorized into three categories [30]:

- Phonological and articulation errors at the phoneme level
- Hypernasality
- Prosodic errors

Existing tools, such as STAR [31], utilize automatic speech recognition (ASR) systems to detect phoneme substitution errors in children with articulation disorders, aiding them in their speech training efforts. Additionally, Case-Based Reasoning (CBR) have been used to analyze specific instances of speech disorders [32]. This approach allows the system to learn from previous cases, enabling it to adapt and make informed decisions based on similar past examples. By applying CBR, computing systems can better understand and respond to the unique needs of each child, improving the effectiveness of speech therapy interventions.

3.4.2.2 Automated Screening Models

Moreover, a model aiming at automatically screen speech development issues in children by identifying PPs in English-speaking children has been developed [33]. The model defines PPs as encompassing deleted, inserted, or substituted phonemes, and their detection allows for categorization of a child's speech into three risk levels (i.e., low, moderate and high).

In their study, they introduced a proof-of-concept system focused on

fronting and gliding PPs, using a small corpus to emphasize the significance of data quality over quantity for optimal model performance. Their pipeline consists of a stack of different models where each model's output serves as input for the subsequent one. Initially, the input speech is fed into a hierarchical neural network (HNN) functioning as the acoustic model to generate probabilities for pronounced phonemes.

Subsequently, these probabilities from the HNN serve as emission probabilities for a constrained Hidden Markov Model (HMM) decoder. Emission probabilities are the likelihoods that the observed data (in this case, the acoustic features of the pronounced phonemes) are generated from a particular hidden state in the HMM. In other words, they indicate the probability of observing a specific phoneme given a certain phonological state. The HMM uses these probabilities to decode the sequence of phonemes by considering both the acoustic input and the underlying phonological processes.

This HMM decoder incorporates a general understanding of common PPs and integrates the most probable error pattern for each target word using a dictionary, collaboratively compiled with the assistance of SLPs. For instance, considering the example of the word 'teeth' (1), their dictionary lists the target phonemes with the key of the dictionary being the index (i.e., the position in the sequence of phoneme) of each successive phoneme (0, 1, 2 etc.) and the value of each key being the target phoneme and acceptable substitutions for each phoneme position based on the age-specific norms (e.g., target phoneme 'TH' at index 2 for the word 'teeth', and acceptable substitutions 'F' or 'T').

1. teeth.dictionary =

0: ['T'],

1: ['IY'],

2: ['TH', 'F', 'T']

Following this, transition probabilities between HMM states are trained

using the dictionary's transitions to adjust the connections. The Viterbi algorithm is then employed to infer the most probable sequence of phonemes. They define three distinct transition weights:

- *Ws* for transitions to the current phoneme state
- *We* for transitions to a phoneme state specified in the dictionary
- *Wu* for transitions to an unexpected phoneme state, where $Wu \neq 0$ to account for the possibility of unexpected transitions

Subsequently, PP detection is performed by aligning the recognized phoneme string from the HMM with the target word using the Needleman-Wunsch algorithm, enabling the detection of regions with phoneme substitutions, deletions, and insertions. A decision tree is utilized to classify the identified patterns as specific PPs. Despite achieving accurate results with limited data, their model is not yet sufficiently robust for deployment as a substitute for SLPs.

3.4.2.3 Maximum Entropy Model: Classification for Diagnosis

Maximum entropy (MaxEnt) models are classifiers that use a set of manually defined constraints to assign probabilities to different outcomes. Unlike HMMs, which struggle with data not present during their training, MaxEnt models are feature-rich classifiers. They combine various heterogeneous features within a probabilistic framework to output the most probable tag or category for a given input. In our case, we aim to classify each child, based on a set of pronounced words, as typical or atypical.

Classification models presented earlier are language-specific, primarily for English. However, MaxEnt models can be easily adapted for any language and can be used to classify a child as typical or with DLD. Furthermore, they have already been extensively utilized in linguistics, demonstrating high accuracy in other classification tasks like part-of-speech tagging [34] and sentiment analysis [35]. They have also found applications in medical diagnosis, where they use patient histories to predict disease likelihood.

Moreover, MaxEnt models have been used in prior studies on phonological development [36], showcasing their compatibility with phonotactic principles and alignment with different phonological theories, including the one used in the present study: Optimality Theory (OT) [3].

In linguistics, their effectiveness is particularly pronounced due to the model's ability to take into consideration the interdependence of features (i.e., each phoneme or word is influenced by its surrounding environment and is not independent of it). The model does its classifying based on certain predefined or induced features, which allows it to capture the complexity of linguistic data more accurately than models assuming feature independence.

Entropy, defined mathematically as shown in (2), is a measure of uncertainty or randomness [37]. This equation calculates the entropy of the probability distribution by summing up the product of each possible value of x (from 0 to infinity) and the natural logarithm of its probability mass function $P_{M_E}(x)$. The negative sign ensures that the result represents entropy, a measure of uncertainty or randomness in the probability distribution.

2. $S = -\sum_{x=0}^{\infty} P_{M_E}(x) \log P_{M_E}(x)$

where:

S represents the entropy of the probability distribution

$P_{M_E}(x)$ denotes the probability mass function (P_{M_F}) of the variable having a particular value x (in our case, the particular value x is the probability of a word form being pronounced by a child)

log represents the natural logarithm

The maximum entropy principle is a method used in probability and statistics to determine the most unbiased probability distribution given a set of known constraints. When faced with several possible probability distributions that could describe a system, the maximum entropy principle chooses the one with the largest entropy.

A distribution with high entropy is more spread out and less certain, meaning it makes fewer assumptions about the unknown aspects of the system. By choosing the distribution with the highest entropy, we are effectively saying that we should not assume any additional information that we do not have. This approach ensures that our model remains as unbiased as possible and only relies on the information that is actually known. In practical terms, this means that the maximum entropy model is the one that is most consistent with the given data while remaining as non-committal as possible about unknowns.

For this reason, MaxEnt models stand out for their proficiency in forecasting future events based on present conditions, particularly in scenarios characterized by complex systems and limited data availability. These models excel when the number of potential system configurations, or data distribution (N) far surpasses the observed data points (K) as represented in (3). In essence, they excel in scenarios where the range of possible system states greatly exceeds the instances observed in real-world data.

3. $N \gg K$

In our study, where data is constrained and does not encompass the entirety of potential language patterns, MaxEnt models guard against over-reliance on existing data (i.e., overfitting) while maintaining the ability to make precise predictions for novel input.

This modeling approach is particularly useful in systems where individual decisions appear disconnected, such as within language patterns where surface-level variability across words, contexts, and individuals may seem disparate. By constraining a few key aspects, MaxEnt models can shed light on a significant portion of the system's complexity, particularly in cases where a comprehensive bottom-up modeling approach is impossible, as is the case with the intricate phonological grammar acquisition process for all (Dutch-)speaking children. For these reasons, we

use a MaxEnt model to classify the children in our study as TD or with DLD.

MaxEnt Model Principles

As stated above, MaxEnt models are classifiers. Classification involves analyzing individual observations (here, pronounced words), extracting relevant features that characterize each observation (here, which phonological processes are used in each pronounced word), and then assigning it to one of several distinct categories based on these features (here, TD or DLD). As such, the output of the MaxEnt is a probability distribution $P_{M_E}(x)$ which must meet three key requirements:

1. It must satisfy a limited number of constraints (in our case, the constraints are defined by the Optimality Theory framework).

2. The probability distribution must be the distribution with the maximum entropy of all distributions that satisfies the defined constraints.

3. The probability distribution must adhere to a mandatory normalization constraint, ensuring that the sum of probabilities for all possible outcomes equals to 1, as expressed in (4).

4. $\sum P_{M_E}(x) = 1$

In summary, MaxEnt models operate within the constraints defined by the data, utilizing a rich set of features to make unbiased predictions. By maximizing entropy within these constraints, the model effectively captures the uncertainty and complexity inherent in the data, thereby facilitating accurate classification tasks.

Mechanisms of MaxEnt Models

MaxEnt models, also known as exponential or log-linear classifiers, op-

erate by extracting a set of features (defined by the modeler) from the input data, which are then combined linearly with weights determined during the model training process. These models utilize a probabilistic framework, with the sum of these weighted features serving as the exponent of a normalization constraint, ensuring that the probabilities sum to 1. In our specific application, which focuses on children's pronunciation of words, we aim to extract relevant features indicating whether specific phonological processes, such as final consonant deletion, are present in the children's speech.

The probability of a particular class being correct given an input x is calculated using the formula depicted in (5).

5. $p(d \mid x) = \frac{1}{Z} \exp\left(\sum_i w_i f_i\right)$

where:

$p(d \mid x)$ represents the probability of a class (decision) being correct given an input x

exp is the exponential

Z serves as a normalizing constant

w_i refers to the weight associated with each feature f_i in the model.

f_i represents the features extracted from the input data x

To better understand this equation, several steps are needed. The mathematical demonstration is presented below.

Features within MaxEnt Models

As modelers, we choose arbitrarily the features functions we want to reflect the characteristics of the problem domain as faithfully as possible. Each feature, also called constraint in our case, is binary, indicating whether it is present in the input or not. Formally, the feature function $f_{\text{cp},y'}(x, y)$ is

defined in (6).

$$6\ f_{\mathrm{cp},y'}(x,y) = \begin{cases} 1 & \text{if } y = y' \text{ and } \mathrm{cp}(x) = \text{True} \\ 0 & \text{otherwise} \end{cases}$$

where:

cp (contextual predicate) corresponds to a given constraint, cp maps a pair of outcome y and a context x to true or false.

An example is given in table 3.1 for our particular case. The input word represents the adult form (i.e., target word transcription), while the output indicates the pronunciation transcription of the word by a child, which are also described in terms of violations of constraints derived from OT.

Each word pair in the table identifies a violation of a defined constraint by the pronounced word. For instance, the constraint '*Coda' signifies the prohibition of codas, where a violation (i.e., presence of a coda) is marked as 1, while adherence to the constraint (i.e., no coda) is represented by 0.

Additionally, the constraint 'MAX' stipulates that the output should closely resemble the input, specifically by prohibiting the deletion of segments present in the input. In our analysis, we focus on the PP 'final consonant deletion'. Since the other PPs we examine involve only substitutions and not deletions, we use MAX in its general form. However, if our analysis included the deletion of different segments in various positions, MAX could be reformulated to address specific deletions, such as prohibiting the deletion of word-initial or syllable-initial segments, or the deletion of vowels.

In the first row, a violation occurs as the child's pronunciation (tO) differs from the input target form (tOt), thus resulting in a violation of this constraint, indicated by 1.

Each feature, denoted as $f_i(x)$, is associated with a weight $w_i(d)$ for each class (or decision), with a weight of 0 when the feature is not present. This weight, determined during training (as it will be explained in the next sub

Input	Output	*Coda	MAX
tOt	tO	0	1
tOt	tOt	1	0

Table 3.1: MaxEnt table input

section), reflects the contribution of the feature to the classification decision. Predicting a decision (d) for a given input x involves evaluating whether each feature is present or not and multiplying it by its associated weight as shown formally in (7).

7. $f_i(x) * w_i(d)$

The subsequent steps involve calculating the numerator and denominator of the probability distribution.

The numerator represents the total weight of features for each class (i.e., one numerator per class), obtained by summing the exponentiated weights of all present features. More formally, for N features, the total of each result of equation (7) is added and takes the exponent of summation to get a numerator determining the weight for each class (8).

8. $numerator_d = exp(\sum_{i=1}^{N} f_i(x) * w_i(d))$

Then, the denominator is obtained by summing the numerators (i.e., the total weights of all classes) (9).

9. $denominator = \sum_{d'} exp(\sum_{i=1}^{N} f_i(x) * w_i(d'))$

Finally, the probability of a class given an input is calculated as the ratio of the numerator of the given class to the denominator (10) which corresponds to equation (5), retranscribed below,, which represents the probabilistic formulation used in MaxEnt models, often referred to as the softmax function.

10 $p(d \mid x) = \frac{numerator_d}{denominator}$

5. $p(d \mid x) = \frac{1}{Z} exp(\sum w_i f_i)$

where:

$p(d \mid x)$ is the probability of class d given an input x

Z (the normalization constant or partition function) ensures that the probabilities sum to 1 over all possible classes d'

To clarify, equation (5) defines $p(d \mid x)$ as the exponentiated weighted sum of features divided by Z, where Z ensures that the probabilities across all classes sum to 1. In practical terms, Z the denominator in equation (18)) is the sum of all exponentiated scores across all classes.

Equation (10) shows how $p(d \mid x)$ is computed for a specific class d using the numerator and denominator defined earlier. The numerator $numerator_d$ represents the likelihood of the class d given the input x while the denominator *denominator* ensures that the probabilities are properly normalized across all possible classes.

Therefore, equation (10) directly implements the softmax function described in equation (5), where Z is the denominator in equation (10). The numerator $numerator_d$ is the specific term for the class d being considered, and *denominator* ensures that $p(d \mid x)$ is a valid probability distribution by summing over all possible classes d'.

This formula enables the model to assign probabilities to each class based on the input features, with the highest probability determining the classification decision.

MaxEnt models training

Training a MaxEnt model involves three essential components:

- Training Data: A set of training data is required, comprising different words pronounced by TD children only.
- Manually Defined Features: Features must be manually defined to capture the relevant phonological processes observed in the children's speech. These features help identify which phonological processes and their frequencies are associated with different types of developing children.
- Parameter Estimation Function: A function is needed to estimate the parameters of the model and assign appropriate weights to each feature to optimize the model's output.

The training process aims to find real-value weights for each feature that maximize the model's log likelihood. Each weight must reflect its importance in determining the classification outcome. This is depicted by the following formula (11).

11. $L(p) = \sum_{x,y} \tilde{p}(x,y) \cdot log p(y \mid x)$

where:

$L(p)$ represents the log likelihood function of the MaxEnt model

$p(y \mid x))$ denotes the probability of the output y given the input x

$\tilde{p}(x,y)$ represents the empirical distribution of outputs y given inputs x observed in the training data

x represents the input data, in our case, the features extracted from the adult pronunciation of a word

y represents the output data, in our case, the different children's pronunciations of the same word

4. Research Questions

Aim of the research: The aim of this project is to use phonetic transcription along with computational techniques to analyze some phonological processes exhibited by Dutch children, providing a detailed summary of these processes for each child's speech data. Additionally, the project aims to develop a classification model to categorize children as typically developing or having developmental language disorder based on these analyzed phonological patterns.

Research Question 1: How accurately can phonological processes be modeled from phonetic transcriptions of Dutch children's speech?

Research Question 2: Can the phonological processes modeled in this research distinguish typically developing children and those with developmental language disorder in our data?

Research Question 3: How effective is a classifier trained on modeled phonological processes in distinguishing between typical children and those with developmental language disorder?

Research Question 4: How reliable are the phonological process models and classifiers across different age groups of Dutch children?

5. Methods

5.1 Data

Importantly, during the initial stages of development, children's language profiles can diverge quite significantly (in terms of advancement) [1], which makes it difficult to distinguish typical from atypical development. This underscores the importance of longitudinal studies to track developmental trajectories and identify potential impairments through the development of the child. Therefore, in collaboration with Auris this work makes use of one dataset gathered by Dr. Everaert on a non-word repetition task (NWRT) [2]. The dataset consists of recordings from children aged 3;0 to 6;3 (years; months), including both typically developing (TD) children and children with developmental language disorder (DLD).

It encompasses audio recordings, manual pseudo-phonetic transcriptions of the repeated words, the intended target words, the child's diagnosis (TD or DLD), gender, and age at the time of recording. The dataset features 30 pseudowords, as well as a words ranging in length from one to five syllables. Details of the target words are provided in Appendix B, Table B.1[1].

It is important to note that Dutch does not typically have five-syllable words, except in compounds or derived words. Furthermore, some of the phoneme combinations present in the target words are not usual sound combinations in Dutch. However, the NWRT is designed to assess phonological memory in children. This means that the task is intended to evaluate how well children can repeat unfamiliar sound sequences, which is a measure of their phonological memory and processing skills, rather than their

[1]The pseudo-phonetic transcription of the NWRT used in this research, including the target words and the pronounced words by the children are available at https://drive.google.com/drive/folders/121uOHP0TUggvkIudvwUrSkwtfkxRc8Zz?usp=sharing

ability to pronounce familiar words in their native language.

Given this focus, the dataset was not originally designed for phonological analysis, which aims to analyze specific phonological processes and patterns in children's speech. Our research repurposes this dataset to provide insights into phonological development and disorders by examining the errors and variations in the children's pronunciations of these non-words.

Pronounced words were collected from 50 children with DLD and 62 TD children. Despite its modest size, this dataset serves as a valuable resource for the initial testing of the algorithms proposed in this study.

5.2 Phonetic Transcription

The initial phase of this research depends on obtaining a reliable broad phonetic transcription from the audio recordings provided by Auris. While other researchers previously utilized a combination of HNN-HMM (Hierarchic Neural Networks- Hidden Markov Model) for this purpose [33], their methodology has become outdated with advancements in state-of-the-art models.

A variant of Wav2Vec [38] offers the capability of phonemic transcription, known as Wav2Vec to Phonemes. This model has already undergone fine-tuning for Dutch phoneme recognition[2] for Dutch adult speakers.

However, the phonetic characteristics of child speech differ from those of adult-directed speech [39]. Therefore, comparisons have been conducted between manual phonetic transcription and automatic transcription provided by Wav2Vec on PhonBank files containing Dutch CHILDES data in phonetic transcription. The Character Error Rate (CER) was utilized to assess transcription accuracy. It is noteworthy that PhonBank only contains manual phonetic transcription of children under 2;07, hence the tests were

[2]https://huggingface.co/Clementapa/wav2vec2-base-960h-phoneme-reco-dutch

conducted only on this data. However, considering our focus on children aged 3;0 to 6;2, it is important to acknowledge that older children may exhibit phonetic features closer to adult speech, but no comparisons could be made. The tests revealed a CER of approximately 0.60, indicating that only 40% of the transcription was correct, which is insufficient for us to use this method.

Current technologies do not offer sufficiently accurate phonetic transcription of child speech due to a lack of training data, making fine-tuning impractical for this model. Therefore, alternative methodologies were considered to obtain accurate phonetic transcriptions for input into the models intended to be developed in this study. However, once accurate phonetic transcriptions from child speech audio will become feasible, the models created here will be ready to take speech audio as input.

Other methods explored include grapheme to phoneme conversion, which are used in speech synthesis systems such as eSpeak[3], given the inability to directly transcribe audio, and the already orthographically transcribed data used in this research. Speech synthesizers like eSpeak take the language name and the orthographic transcription as input and produce phonetic transcriptions following the phonological rules of the specified language. Though initially prioritizing direct audio analysis, speech synthesis would be utilized if proven more feasible.

However, after testing eSpeak, this method's reliance on phonological rules resulted in inaccurate transcriptions, particularly with pseudo-words, as it attempted to conform to adult speech patterns, yielding a CER similar to that of Wav2Vec.

Therefore, seeing that the various approaches to automatic transcription of the children's speech failed, a custom algorithm has been developed to use the pseudo-phonetic transcription made by SLPs using KAT to generate reliable phonetic transcriptions for the models.

Taking as example (12), four non-words from our dataset.

[3]https://espeak.sourceforge.net/

12. [‘Keepon’, ‘Sietaalon’, ‘Peelaanot’, ‘Liepoetaan’]

The custom algorithm works as follow;

1. Using `Indicsyllabifier`[4] along with manual adjustments of this Python module, we separate in syllables the pseudo-phonetic transcription of the words from the data:
   - ‘kee’, ‘pon’
   - ‘sie’, ‘taa’, ‘lon’
   - ‘pee’, ‘laa’, ‘not’
   - ‘lie’, ‘poe’, ‘taan’
2. Looking for the vowel or the diphthong in each syllable, we split each syllable into what is before the vowel or the diphthong (the onset), the vowel or the diphthong (the nucleus) and what is after (the coda), and we make sure to keep three elements in each syllable, even if some are empty, for alignment in further processing:
   - [‘k’, ‘ee’,‘ ’], [‘p’, ‘o’, ‘n’],
   - [‘s’, ‘ie’,‘ ’], [‘t’, ‘aa’, ‘ ’], [‘l’, ‘o’, ‘n’],
   - [‘p’, ‘ee’, ‘ ’], [‘l’, ‘aa’, ‘ ’], [‘n’, ‘o’, ‘t’],
   - [‘l’, ‘ie’, ‘ ’], [‘p’, ‘oe’, ‘ ’], [‘t’, ‘aa’, ‘n’]
3. As the input is a pseudo-phonetic transcription, each character has only one correspondence in IPA, making the transformation in IPA feasible. We also keep the split into syllables and the split of syllables into onset, nucleus and coda, and also adding a representation of the entire word into phonetics:
   - ‘kepɔn’: [[‘k’, ‘e’, ‘’], [‘p’, ‘ɔ’, ‘n’]],

[4]https://silpa.readthedocs.io/projects/indicsyllabifier/en/latest/

- 'sitalɔn': [['s', 'i', ''], ['t', 'a', ''], ['l', 'ɔ', 'n']],
- 'pelanɔt': [['p', 'e', , ''], ['l', 'a', ''], ['n', 'ɔ', 't']],
- 'liputan': [['l', 'i', ''], ['p', 'u',„ ''], ['t', 'a', 'n']]

5.3 Phonological Processes

This sub-section describes the phonological processes (PPs) that will be used in our models.

Beer gives a list of the PPs found in Dutch children [13]. As stated in section 3.3, she categorizes them as a) the ones found in normal phonological development, which have the strongest negative effect on intelligibility[5], b) unusual processes, appearing in normal development but not so often and c) the ones appearing seldomly in typical development.

As this research is a pilot study, we focus on four very common PPs, and will use our models on them. Their description along with examples taken from Beer's research and the likely age of disappearance are presented in Table 5.1.

It is important to note that other phonological processes may also take place in the studied words (but also in the given examples in Table 5.1). For instance, in the example 'blas@' vs. 'pat', multiple PPs might be involved. In our analysis, we focus only on a selected few PPs. This means that other processes present in the children's speech are not accounted for in our models.

By ignoring these additional processes, our results may provide an incomplete picture of the children's phonological development. Consequently, our interpretations will be limited to the specific PPs we have chosen to study. Future research should aim to include a broader range of PPs and study them in interaction to provide a more comprehensive analysis of

[5] As Beers notes, these processes significantly affect intelligibility in children with DLD as they progress through later stages of acquisition. This impact is exacerbated by the uneven development among phonology, morphosyntax, and vocabulary in these children.

phonological development and disorders.

Substitution **One sound is substituted for another sound in a systematic way**				
Process	Description	Target	Realization	Likely age of elimination
Fronting	A sound produced in the back of the mouth (ie. velar) is replaced by a sound in the front of the mouth (ie. alveolar)	hOŋ@r (honger)	hOn@r	- 4yo
Stopping	A fricative is replaced by a stop	blas@ (blazen)	pat	- 3yo for /f,s/ - 4yo for /z,v/
Gliding	A liquid is replaced by a glide	ram@ (ramen)	wam@	- 6 to 7yo
Syllabic structure **Sound changes that affect the syllable structure of a word**				
Process	Description	Target	Realization	Likely age of elimination
Final consonant deletion	Deletion of the final consonant of the word	tOt (tot)	tO	- 3yo

Figure 5.1: Phonological Processes that the present study focuses on

5.4 Automatic Detection of Phonological Processes

In this sub-section, we elaborate on the methodology employed in constructing the two models within this pipeline. The first model focuses on automatically detecting phonological transformations, while the second

model utilizes this output for classifying children as typically developing (TD) or with developmental language disorder (DLD), encompassing their respective implementations and considerations.

5.4.1 Levenshtein Distance and Breadth First Search Algorithm

The Levenshtein Distance (LD) combined with the Breadth First Search (BFS) algorithm (denoted as LD-BFS) presents the initial step for investigation.

This approach aims to identify phonological processes used by children by enumerating the segmental discrepancies between the target word (i.e., the word the children are supposed to pronounce) and the word the children actually pronounced. These discrepancies inform a metric that captures the phonological 'distance' between the two forms. The LD creates a matrix with all possible insertions, substitutions and deletions to go from the pronounced word to the target word, and the BFS searches for the shortest way to go from the former to the latter and returns the edits (i.e., insertions, substitutions and deletions) used during this shortest path.

To explain the LD-BFS implementation, we can use one reference pseudoword from our dataset represented in pseudo-phonetic (13.a) and in IPA (13.b), along with a deviation by a child, as shown in (14.a; pseudo-phonetic transcription) and (14.b; IPA transcription) respectively.

13.a sietaalon

13.b sitalɔn

14.a tietaaong

14.b titaɔŋ

Constructing a LD matrix for the example 'sitalɔn' vs 'titaɔŋ' yields

(15). The numbers in the matrix represent the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform one part of a word into another. Each cell in the matrix indicates the edit distance between the corresponding segments of the target word and the child's pronunciation up to that point. The upper left corner starts at 0, as no edit is needed initially. The process of filling in the matrix is done from the beginning (left) to the end (right) of the words.

For example, the cell [2,2] has a value of 1, indicating that one edit is required to transform the first character of the child's pronunciation into the first character of the target word. Specifically, 't' is substituted with 's'. Moving on to cell [3,3], it also has a value of 1. This is because the 'i' in both positions matches, so no additional edit is needed, but the previous edit (substituting 't' with 's') is carried forward. Thus, the process involves mapping each segment of the child's pronunciation onto the corresponding segment of the adult target word incrementally from left to right, accounting for the minimal edits required at each step.

(15) Distance Matrix

		s	*i*	*t*	*a*	*l*	ɔ	*n*
	[0,	1,	2,	3,	4,	5,	6,	7]
t	[1,	1,	2,	2,	3,	4,	5,	6]
i	[2,	2,	1,	2,	3,	4,	5,	6]
t	[3,	3,	2,	1,	2,	3,	4,	5]
a	[4,	4,	3,	2,	1,	2,	3,	4]
ɔ	[5,	5,	4,	3,	2,	2,	2,	3]
ŋ	[6,	6,	5,	4,	3,	3,	3,	3]

Using BFS enables us to identify the shortest path (16) in the LD matrix (15), indicated by the starred cells and the required edits to transition from the reference to the pronounced word (17). A cell becomes starred if it lies

on the path representing the minimum number of edits needed to transform the child's pronunciation into the target word. This path is determined by tracing back through the matrix from the bottom-right cell (which shows the total edit distance) to the top-left cell, always moving to the neighboring cell that contributed to the current cell's edit distance (i.e., the minimum edit distance from an insertion, deletion, or substitution). The starred cells thus represent the specific sequence of edits required to make this transformation.

In this approach, each shift from one column to the next column on the right side signifies an insertion, a shift from one row to another row below signifies a deletion, and diagonal movement indicates either no edit (if the number remains the same) or a substitution (if it increases).

(16) Path

		s	*i*	*t*	*a*	*l*	ɔ	*n*
	[*,	1,	2,	3,	4,	5,	6,	7]
t	[1,	*,	2,	2,	3,	4,	5,	6]
i	[2,	2,	*,	2,	3,	4,	5,	6]
t	[3,	3,	2,	*,	2,	3,	4,	5]
a	[4,	4,	3,	2,	*,	*,	3,	4]
ɔ	[5,	5,	4,	3,	2,	2,	*,	3]
ŋ	[6,	6,	5,	4,	3,	3,	3,	*]

(17) Edits

- ('substitution', 't', 's')
- ('no edit', 'i', '')
- ('no edit', 't', '')
- ('no edit', 'a', '')
- ('insertion', '', 'l')

- (‘no edit’, ‘ɔ’, ‘’)
- (‘substitution’, ‘ŋ’, ‘n’)

Syllables are fundamental building blocks in spoken language and play a critical role in various aspects of phonological processing. By focusing on syllables, we can achieve a more granular and accurate comparison, which is particularly useful in detecting and analyzing pronunciation errors and phonological patterns in children’s speech. Therefore, as syllables hold greater linguistic significance for analysis rather than the analysis of entire words, matrices are not computed for the whole words but for each syllable, comparing distinct units such as onset, nucleus, and coda.

Given that words are already segmented into meaningful sub-units during phonetic transcription, as explained in section 5.2, the process described above can be iterated for these sub-units.

In our example, it give three matrices, and thus three shortest paths (18).

18.

		s	*i*
	*	1	2	3
t	1	*	2	3
i	2	2	*	2
	3	3	2	*

Edits: (‘substitution’, ‘t’, ‘s’) (‘no edit’, ‘i’, ‘’) (‘no edit’, ‘ ’, ‘’)

		t	*a*
	*	1	2	3
t	1	*	1	2
a	2	1	*	1
	3	2	1	*

Edits: (‘no edit’, ‘t’, ‘’) (‘no edit’, ‘a’, ‘’) (‘no edit’, ‘ ’, ‘’)

		l	ɔ	*n*
	*	*	2	3
ɔ	1	1	*	2
ŋ	2	2	2	*

Edits: (‘insertion’, ‘’, ‘l’) (‘no edit’, ‘ɔ’, ‘’) (‘substitution’, ‘ŋ’, ‘n’)

In this approach, each syllable, even if containing empty elements, is divided into three components: onset, nucleus, and coda. This breakdown ensures a precise comparison between corresponding elements in the reference and the child’s pronounced word.

Initially, the syllables in both the reference and the child’s pronounced word are aligned. If the child’s pronunciation exhibits one or more syllables with solely empty elements, it signifies syllable deletion, and the index of the deleted syllable is documented to identify which syllable from the reference word is missing. This information is included in the detailed summary of insertions, deletions, and substitutions returned by the LD-BFS algorithm but is not further processed in the MaxEnt model, as syllable deletion is not part of the phonological processes analyzed in this study.

The resulting output serves as the foundation for aligning these edits with specific phonological processes, such as determining whether a deleted consonant is part of cluster reduction or final consonant deletion, thereby facilitating the final diagnostic analysis.

Each child’s data is structured into a CSV file, illustrating various aspects such as phonetic transcriptions, syllable indices, and phoneme edits. This CSV file, as exemplified in Figure 5.2, presents the edits of each word pronounced, including columns for the phonetic transcription of the reference word, the phonetic transcription of the pronounced word, the index of the syllable of the current phoneme under investigation, the edit type (no edit,

insertion, substitution, deletion), the phoneme in the reference word, and the phoneme in the pronounced word.

Reference	Pronounced	Edit Position	Edit	Reference sound	Pronounced sound
liputan	liputa	sublist 1	deletion		l
liputan	liputa	sublist 1	no edit	i	i
liputan	liputa	sublist 2	substitution	p	b
liputan	liputa	sublist 2	no edit	u	u
liputan	liputa	sublist 3	no edit	t	t
liputan	liputa	sublist 3	no edit	a	a
liputan	liputa	sublist 3	insertion	n	

Figure 5.2: Output of the LD-BFS for the reference word 'liputan'

This CSV file is pivotal for formatting the input data for the subsequent MaxEnt model, which classifies each child based on their phonological characteristics. Moreover, additional CSV files are generated to comprehensively analyze the phonological processes utilized by each child, categorized by age range and diagnosis, akin to the approach undertaken by KAT.

Furthermore, specific CSV files are generated for each age range and diagnosis, listing the phonological processes employed by each child, or indicating an empty list if no processes were used.

5.4.2 MaxEnt Model

This subsection presents the use of a Maximum Entropy (MaxEnt) model as the final phase of the analysis pipeline, integrating phonological analysis as features for classification.

The chosen MaxEnt implementation, developed by the University of Massachusetts Amherst[6], operates efficiently with a simple input structure:

[6]https://websites.umass.edu/hgr/

a list of target adult words (input), corresponding child pronounced words (output), a probability of each pronounced word to be an output, and binary indicators (0 or 1) denoting the satisfaction or violation of manually predefined constraints.

Before explaining this implementation in more details it is important to explain how the constraints are defined and formulated.

Drawing from the Optimality Theory (OT) framework, this research adheres to the prevalent view in phonological acquisition, positing that children's mispronunciations stem from differences in their phonological systems, prioritizing markedness constraints over faithfulness constraints, thereby yielding more unmarked productions. While this perspective guides the definition of constraints in this study, it's noteworthy that alternative viewpoints exist. Some scholars in OT, for instance, argue for the inclusion of performance limitations in their constraint formulation, introducing a separate COST family constraint operating at the articulatory level rather than solely within the internal grammar [40].

Although this pilot study aligns with the predominant OT perspective on phonological acquisition, the MaxEnt model itself remains theory-agnostic, allowing for adaptation to alternative phonological theories or perspectives.

Hence, it's imperative to explain how the phonological processes outlined in Section 5.3 are translated into OT constraints.

5.4.2.1 Transforming phonological processes into OT constraints

For our analysis, we translate the four phonological processes (PPs) described in section 5.3 into OT constraints. The PPs investigated include final consonant deletion, stopping, fronting, and gliding.

Children's speech patterns, as explained by OT, are governed by

markedness constraints that rank higher than faithfulness constraints. Notably, faithfulness constraints encompass three key families:

- Max: Against the deletion of an element present in the input
- Dep: Against the insertion of an element not present in the input
- Ident[feature]: Against the substitution of a feature present in the input by another feature

On the other hand, markedness constraints, while more numerous and less precisely defined, are selectively identified for our study.

Final Consonant Deletion

This process, typically used until approximately 3;0, involves the deletion of the final consonant of a word. Within a child's grammar, the *Coda markedness constraint typically ranks higher than the Max constraint (19), as illustrated in Table 5.3.

The *Coda constraint prohibits consonants in the syllable coda position, reflecting a preference for open syllables (CV structures). This preference is related to markedness, which may be influenced by factors such as articulatory difficulty, frequency of occurrence in the language or late emergence in acquisition. Indeed, as stated before, children firstly use unmarked structures, and the most unmarked syllabic structure is the open syllable CV, which leads to a prohibition of syllables with codas.

19. *Coda $\gg$ Max

Table 5.3 illustrates the hierarchical relationship between constraints in the context of systematic Final Consonant Deletion. It presents two candidate forms, taken from Beers' study [13]: a faithful representation (tOt), perfectly matching the input, and another form (tO) where a final consonant has been deleted.

In this scenario, *Coda is a markedness constraint stipulating the absence of a coda, while MAX is a constraint ensuring the output resembles the input. In the table, *Coda holds a higher rank than MAX. Consequently, the

faithful candidate is eliminated (fatal violation indicated by '!'), as violating the higher-ranked constraint. Despite the unfaithful candidate contravening the lower-ranked MAX constraint, it remains the chosen output

	*Coda	MAX
-> tO		*
tOt	*!	

Figure 5.3: First stages of child grammar development: Systematic Final Consonant Deletion

However, as a child progresses in acquisition, occasional final consonant deletions may still occur, albeit inconsistently across words. In such cases, OT suggests a reevaluation of constraint hierarchy. While the faithful candidate begins to ascend in the hierarchy, it does not yet surpass the markedness constraint. At this stage (as indicated in Table 5.4 by the broken line between the two constraints), both constraints are positioned at the same level (20), allowing for multiple optimal output candidates. This phenomenon underscores the dynamic nature of constraint ranking and reranking during the stages of acquisition, particularly in child development contexts.

20. *Coda, Max

	*Coda	MAX
-> tO		*
-> tOt	*	

Figure 5.4: Later stages of child grammar development: Optional Final Consonant Deletion

	MAX	*Coda
tO	*!	
-> tOt		*

Figure 5.5: Adult grammar: No Final Consonant Deletion

Finally, in adult Dutch language, MAX is ranked above *Coda, as Dutch allows for codas, and the unfaithful candidate /tO/ is eliminated, as shown in table 5.5).

Fronting

A similar phenomenon is observed in the case of fronting, which typically persists until the child reaches the age of 4;0. Here, the markedness constraint *Coronals, which prohibits coronal elements, takes precedence over the faithfulness constraint IDENT[coronal], where the coronal element in the input must be identical in the output. This reflects a preference for anterior sounds as they are more unmarked, compared to posterior sounds like coronals. This hierarchy is illustrated in (21) and exemplified in Table 5.6.

When fronting occurs sporadically or inconsistently, both *Coronals and IDENT[coronal] constraints are placed at the same stratum, as illustrated in (22) and depicted in Table 5.7. This indicates an equal level of significance for both constraints, allowing for the possibility of multiple optimal output candidates.

21. *Coronals ≫ IDENT[coronal]

22. *Coronals, IDENT[coronal]

hOŋ@r	*Coronals	Ident[cor]
-> hOn@r		*
hOŋ@r	*!	

Figure 5.6: Systematic Fronting

	*Coronals	Ident[cor]
-> hOn@r		*
-> hOŋ@r	*	

Figure 5.7: Optional Fronting

Stopping

A similar pattern is observed in the case of stopping, which typically persists until the child reaches the age of 3;0 to 4;0. Here, the markedness constraint *Fricatives, which prohibits fricative elements, is prioritized over the faithfulness constraint IDENT[continuant], where the continuant element in the input must be identical in the output. This reflects a preference for plosives as they are more unmarked, compared to fricatives. This hierarchical relationship is illustrated in (23) and exemplified in Table 5.8.

When stopping occurs inconsistently or sporadically, both *Fricatives and IDENT[continuant] constraints are placed on the same stratum, as depicted in Table 5.9 (24). This indicates an equal level of importance for both constraints, allowing for the possibility of multiple optimal output candidates.

23 *Fricatives ≫ IDENT[continuant]

24. *Fricatives, IDENT[continuant]

	*Fricatives	Ident[cont]
-> pat		*
blas@	*!	

Figure 5.8: Systematic Stopping

	*Fricatives	Ident[cont]
-> pat		*
-> blas@	*	

Figure 5.9: Optional Stopping

Gliding

A similar developmental progression is observed for gliding, typically persisting until the child reaches the age of 6;0 to 7;0. In this scenario, the markedness constraint *Liquids, which prohibits liquid elements, is ranked higher than the faithfulness constraint IDENT[consonant], where the conso-

nant element in the input must remain identical in the output. This reflects a preference for glide sounds as they are more unmarked, compared to liquid sounds which are part of the last phonemes a child acquires. This hierarchical relationship is illustrated in (25) and exemplified in Table 5.10.

When gliding occurs inconsistently or sporadically, both *Liquids and IDENT[consonant] constraints are placed on the same stratum (26), as depicted in Table 5.11. This indicates an equal level of importance for both constraints, allowing for the possibility of multiple optimal output candidates.

25. *Liquids ≫ IDENT[consonant]

26. *Liquids, IDENT[consonant]

	*Liquids	Ident[cons]
-> wam@		*
ram@	*!	

Figure 5.10: Systematic Gliding

	*Liquids	Ident[cons]
-> wam@		*
-> ram@	*	

Figure 5.11: Optional Gliding

For the PPs of fronting, stopping, and gliding, one might wonder how OT ensures that the phoneme in the target word is substituted by a phoneme within the appropriate class determined above. Using gliding as an example, the substitution process at a more abstract phonological level aims to retain as many features of the target sound as possible. Within the OT framework, the replacement of liquids specifically by glides (i.e., w or j) rather than other sounds is due to higher-ranked faithfulness constraints like IDENT[manner], which preserve the manner of articulation of the consonant. This means that liquids, which are approximants, will be

replaced by other approximants (glides), rather than by consonants with different manners of articulation, thereby maintaining as many features of the target word as possible.

In summary, we have demonstrated how each phonological process under study can be translated into constraints with varying rankings based on the age of typically developing children. This process is crucial for understanding the developmental trajectory of phonological acquisition within this framework.

Implementation of constraints into the MaxEnt model

Moving forward, we create three distinct models for different age groups (3-4 years, 4-5 years, and 5-6 years), in which we incorporate the specific constraints into the MaxEnt model.

We use two R scripts developed by the University of Massachusetts Amherst[7] for this process: one for training the model and another for testing it. The training script starts by setting the initial weights to zero and uses an optimization algorithm called Limited-Memory Variable Metric (L-BFGS) to adjust the weights until the model performs well (i.e., until convergence). This algorithm begins with an initial guess for the optimal weights and iteratively improves upon this guess.

The training script employs two types of regularization methods to optimize the model: L1 (Lasso) regularization and L2 (ridge) regularization. Regularization helps determine the weights of each constraint during training.

- **L1 Regularization:** L1 regularization adds a penalty to the model based on the sum of the absolute values of the weights. This helps create simpler models by encouraging some weights to be exactly zero, effectively eliminating some features. This method is robust against outliers and makes the model more interpretable by reducing

[7] https://websites.umass.edu/hgr/

the number of non-zero weights.

- **L2 Regularization:** L2 regularization adds a penalty based on the sum of the squared values of the weights. It helps prevent overfitting by penalizing large weights, distributing the weight values more evenly across features, and promoting smoother models. L2 regularization also improves numerical stability, especially when dealing with features that are highly correlated. L2 plays on the variance of the prior distribution, as a high variance could lead to overfitting, L2 regularization finds a new line which doesn't fit the training data too well by introducting a small amount of bias in the variance, getting a lower variance and better long term prediction.

The main difference between these methods is that L2 regularization can only reduce weights close to zero, while L1 can reduce them to exactly zero, which is useful for eliminating irrelevant features.

In our case, since all constraints are relevant, L2 regularization is more suitable than L1. By default in our implementation, L2 regularization uses a variance of 1000, which allows for more flexibility in the weights.

5.4.2.2 Split in training/ test and k-fold cross validation

As our dataset is pretty small, we decided to use k-fold cross validation, a method commonly used in machine learning to ensure more reliable results. Here's how it works:

1. **Splitting the Dataset:** We split the dataset into k groups (or folds).
2. **Training and Testing:** We train the model k times, each time using a different fold as the test set and the remaining folds as the training set.
3. **Averaging Results:** The final performance is the average of the k test results. This method reduces the risk of our results being due to chance, providing a more accurate estimate of how the model performs on unseen data.

Our approach implementing the k-fold cross validation works as follow:

1. **Initial Split:** We first split the data into different age groups and types of children, as shown in Table 5.12. We reserve about 10% of each group as the test set and use the rest for training.
2. **Choosing k:** We use k=9 for all age groups. This means we split each group into 9 folds, ensuring that each child's data is in the test set only once.
3. **Training and Testing Process:** For each fold, we train the model on the training set and test it on the test set. This process is repeated for each fold, and we use the test results to evaluate the model.

	Typically Developing	Developmental Language Disorder
3;0 - 4;0	16/2	12/1
4;1 - 5;0	16/1	21/2
5;1 - 6;2	25/3	15/1

Figure 5.12: Split of the children per age and diagnosis, and split in training (first number) and test (second number)

5.4.2.3 Formatting the input files

We create an algorithm to process the csv data files (exemplified in Table 5.2) outputted by the previous algorithm. It works as follow:

- **Processing Each Word:** For each word pronounced by a given child, the algorithm takes as input the CSV file produced by the LD-BFS, as shown in Figure 5.2. It examines each phoneme in the target word (one per line) to determine if it has been edited (insertion, deletion, substitution) and whether the edit corresponds to one of the phonological processes (PPs) studied in this research by analyzing the edited phoneme. In the example given in Table 5.2, the reference sound 'p' is substituted by the phoneme 'b', which does not correspond to any of the PPs analyzed here, so no output is generated for this particular substitution.
  - If a change is due to a phonological process studied here, it records a violation of the faithfulness constraint and no violation

of the markedness constraint.

- If the process doesn't apply to the given word, it records no violation.
- If there is no edit but one of the studied phonological process could have applied, it records a violation for the markedness constraint and no violation for the faithfulness constraint.

- **Output Files:** The algorithm generates a text file for each child containing:
  - The target word (input)
  - The pronounced word (output)
  - The constraints (violations or non-violations)

5.4.2.4 Calculate the probability of each word in the input files

To determine the probability of a word being the output of a given input word, several steps are undertaken:

- **Merging Training Files:** We combine all training files into one.
- **Counting Occurrences:** For each input target word, we count how many times each output pronounced word appears.
- **Probability Formula**: The probability of a word being an output of a given input word $P(y_i \mid x_i)$ is calculated as (27).

  27. $P(y_i \mid x_i) = \frac{y_i}{\text{count}(x_i)}$

  where:

  y_i is the pronounced word

  x_i is the target word

It is essential to compare each output to other plausible unseen candidates (i.e., unseen possible combinations of the values for the constraints in each word[8]). This is because the set of candidates represents all possible

[8]The list of the unseen possible candidates created for this research are available at https://drive.google.com/drive/folders/121uOHP0TUggvkIudvwUrSkwtfkxRc8Zz?usp=sharing

pronunciations. We can visualize this as an N-dimensional space, where N is the number of processes considered. For example, if we have three processes, each with two states (0 or 1), the space forms a 2x2x2 cube, where each point represents a different pronunciation variant. Some points may merge if a process is not applicable (e.g., final consonant deletion when there is no final consonant).

Every input should have some probability defined for each combination of processes that are considered. If we don't account for all plausible candidates, we risk creating a biased model by assuming certain probabilities are impossible, rather than simply not observed in our data.

The set of candidates must not be dictated by what is attested for any specific word, but by the processes attested within the entire dataset, and their possible combination so that the probability distribution for different words can be compared to one another.

Moreover, it is important to always compare candidates with the fully faithful candidates (for which $x_i = y_i$) if they are not present in the observed data.

Therefore:

- For the training set:
  - We include all plausible and fully faithful (but unattested candidates in our training set) with a probability of 0. This ensures they are considered possible but not probable based on our data. The MaxEnt model will avoid overfitting by selecting the probability distribution with the highest entropy. It is crucial to assign these unattested candidates a probability of 0 during training because they are not actually observed in our data, and assigning any non-zero probability without strong justification would artificially inflate their likelihood, leading to inaccurate modeling.
- For the test set:
  - We follow a similar approach by integrating all plausible and

fully faithful unattested candidates in our dataset.

– Instead of assigning a probability of 0 to unseen candidates, we use Laplace smoothing. This method assigns a small, non-zero probability to ensure no possible pronunciation is completely ignored. It is essential to use smoothing in the test set because it cannot have zero probabilities; otherwise, the model would unfairly penalize unseen but plausible pronunciations, potentially leading to biased or inaccurate performance metrics.

5.4.2.5 Laplace smoothing

Laplace smoothing (28) handles the issue of 0 probabilities in probabilistic models by adding a small constant (usually $\alpha = 1$) to each count. This way, non-observed but plausible words receive a non-zero probability, ensuring a more realistic and flexible model.

28. $P_{\text{Laplace}}(w) = \frac{C(w)+\alpha}{N+\alpha \times |V|}$

where:

$C(w)$ is the count of the word in the data

α is the smoothing parameter (usually 1)

N is the total number of words observed

$|V|$ represents the size of the vocabulary, which is the number of distinct types of uttered words (as opposed to the total number of word tokens)

Finally, we split the merged test files back into individual files to test the model on each child separately. This approach ensures that our model is tested fairly and can generalize well to new data.

5.4.2.6 Evaluation of the performance on the test sets

Our chosen implementation does not directly classify children as TD or with DLD. Instead it returns a probability for each word in the test set, indicating

how likely it is given the training data. We use an outlier detection approach by training the model exclusively on TD children and then testing it on both TD and DLD children.

For each child in the test set, we start by removing any plausible unseen candidates to focus only on the probabilities of the words actually pronounced by the child. We then calculate the mean and median probability of all the words pronounced by each child, repeating this process for all test files after completing k-fold cross-validation.

Using the real diagnosis of each child (TD or DLD), we set a random threshold to classify children based on the probabilities. If a child's mean or median probability is above this threshold, they are classified as TD; if below, they are classified as DLD. We evaluate the model's performance by assessing its classification outcomes in comparison to the gold standard (i.e., traditional clinical evaluation):

- **True Positives (TP):** Instances where the model correctly identifies children with DLD.
- **False Positives (FP):** Instances where the model incorrectly identifies TD children as having DLD.
- **False Negatives (FN):** Instances where the model incorrectly identifies children with DLD as TD.
- **True Negatives (TN):** Instances where the model correctly identifies TD children as TD.

We repeat the classification process with different thresholds. For each threshold, we record the true positive rate (TPR) defined as in (29) and false positive rate (FPR) defined as in (30).

29. $\text{TPR} = \frac{\text{TP}}{\text{TP}+\text{FN}}$

30. $\text{FPR} = \frac{\text{FP}}{\text{FP}+\text{TN}}$

These rates are then plotted on a Receiver Operating Characteristic (ROC) curve, which illustrates how well a classification model performs across different thresholds. By examining the ROC curve, we can identify

the optimal threshold that provides the best classification results.

The AUC (Area Under the curve) algorithm calculates the total area under the ROC curve, automatically identifying the best threshold. AUC offers an overall measure of model performance across all thresholds. It can be interpreted as the likelihood that the model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. The AUC value ranges from 0 to 1, where 0 indicates a completely incorrect model, and 1 indicates a perfectly accurate model.

This method ensures a thorough evaluation of the model's ability to differentiate between typical and atypical children, leading to the identification of the most effective classification threshold.

6. Results

6.1 LD-BFS output results

The LD-BFS provides detailed phoneme-by-phoneme analysis for each word, identifying instances of insertion, substitution, and deletion, as exemplified in Table 5.2 in the Methods section. Additionally, a comprehensive CSV file is generated, detailing the phonological processes (PP) studied in this research that were used by each child, the phonemes substituted or deleted, and the corresponding words, as shown in Table 6.1.

A final CSV file aggregates the mean frequency of each phonological process by age and diagnosis, offering valuable insights for diagnostic evaluation. This file also includes one metric commonly used by speech and language pathologists (SLPs):

- **Percentage of Consonants Correct (PCC)**: This metric provides an ordinal severity scale indicating the level of disability, intelligibility, and handicap for consonant sounds.

Other metrics exist, but as our research focuses only on individual consonant phonemes, only this measure is relevant to what has been studied here. However, in later work, if more class of sounds or combination of sounds (e.g., clusters) are analyzed, other measures will be relevant to compute.

These metrics enable the identification of a child's stage of acquisition, the evaluation of proximity to target words, and the assessment of word complexity beyond the segment level [41]. Using a single measure does not effectively differentiate between typically developing (TD) children and those with developmental language disorder (DLD). However, combining multiple measures in a regression model provides evidence of disorder [42].

ID Child	Final Consonant Deletion	Fronting	Stopping	Gliding
1111_47_2_1	[]	[]	[]	l -> j Pronounced Word: jylemɪk Reference Word: jyjemʏk
1162_37_1_1	k -> / Pronounced Word: mœypɔt Reference Word: nœypɔk	k -> t Pronounced Word: jet Reference Word: jik	s -> t Pronounced Word: titaːlɔm Reference Word: sitalɔn	[]
1123_36_1_1	n -> / Pronounced Word: nimputa Reference Word: liputaːn	[]	[]	[]
1107_44_1_1	[]	k -> t Pronounced Word: nœypɔt Reference Word: nœypɔk	[]	[]

Figure 6.1: Phonological processes used by four typically developing children aged between 37 and 47 months old

Summaries for our data for TD children and DLD children are shown in Table 6.2, and in Table 6.3.

Our data comes from a non-word repetition task (NWRT). Research on DLD reveals that NWRTs serve as markers for impairment, as children with DLD typically perform poorly on such tasks [16]. Consequently, poor NWRT performance indicates abnormal constraints on word learning, as these tasks typically involve perceiving, storing, and (re)producing non-words. NWRTs are useful for highlighting differences in phonotactic probabilities between children with DLD and TD children.

Age Range	Number of children	Pronounced	Final Cons Del	Fronting	Stopping	Gliding	PCC
36-48	18	0,562	0,009	0,078	0,104	0,031	0,71
49-60	17	0,825	0,007	0,009	0,008	0,047	0,793
61-75	26	0,857	0,01	0	0,002	0,016	0,827

Figure 6.2: Mean Percentage of phonological processes used by typical children from our data, separated per age range in month

Age Range	Number of children	Pronounced	Final Cons Del	Fronting	Stopping	Gliding	PCC
36-48	13	0,36	0,005	0,165	0	0,05	0,612
49-60	21	0,437	0,007	0,07	0,15	0,047	0,641
61-75	16	0,472	0,01	0,051	0,114	0,157	0,744

Figure 6.3: Mean Percentage of phonological processes used by atypical children from our data, separated per age range in month

Statistical analyses were conducted to determine if there is an effect of diagnosis (i.e., typical or atypical) on the frequency of use of the phonological processes studied and the percentage of pseudowords presented that the children repeated (either correctly or incorrectly). The statistical analyses provide empirical evidence to support the foundational assumptions of this project. They help confirm that the phonological processes and pronunciation patterns under focused are indeed different between TD and DLD children in our data. This confirmation is crucial for ensuring that the MaxEnt models are trained on relevant and significant features that truly differentiate TD from children with DLD.

We first look at the normality of the distribution of our data. For this purpose, we split our results by diagnosis, and we perform a Shapiro-Wilk test on each group to observe if the distribution of each group is normal. As can be seen in Appendix C, Table C.1, the result of the Shapiro Wilk p (i.e., the p returned by the test) for each group is always significant, signifying that the null hypothesis (i.e., our data is normally distributed) must be rejected. Therefore our data is not normal, except for the percentage of pronounced words for the children with DLD. But as it is the only group with a normal distribution, we decide to perform non-parametric tests on our data.

A Mann-Whitney test is performed, and it reveals a strong effect of di-

agnosis on the percentage of pseudowords presented that the children repeated across all age ranges ($p = 0.008$ for 3-4 years old, $p < 0.001$ for 4-5 and 5-6 years old) with TD children repeating a higher percentage of pseudowords compared to children with DLD. No significant effect was found for final consonant deletion and stopping, possibly because these phonological processes are typically abandoned before age 3;0, and the children in our study are older.

The test showed a significant effect of diagnosis for fronting in the 5-6 years old group ($p = 0.02$), suggesting that this process, which is usually resolved by age 4;0, can highlight differences between the two groups. Similarly, gliding showed an effect in the oldest group ($p = 0.01$), with this process typically used until around 6:0 to 7;0 years old. Specifically, the DLD group exhibited more instances of these processes compared to the TD group. However, due to the small sample size, no definitive conclusions can be drawn from these visualizations. In summary, the Mann-Whitney test revealed few significant effects, indicating the need for more data to better analyze the differences between TD children and those with DLD.

We also did data visualizations to compare the use of phonological processes (PPs) across ages and groups. These visualizations, presented in Appendix C (Figures C.2, C.3, C.4, C.6 and C.5) indicate trends in the use of PPs. For fronting, both TD children and those with DLD showed a decreasing trend. For final consonant deletion, gliding, and stopping, TD children exhibited these processes less frequently as they aged, whereas children with DLD showed an increasing trend.

Since the increased use of these processes might be correlated with the increased number of pronounced words, we performed a Spearman correlation analysis to investigate these potential relationships.

When analyzing all age ranges together, the Spearman correlation matrix in Table C.7 in Appendix C, revealed a moderate positive correlation between age and the percentage of pronounced words ($r = 0.36$, $p < 0.001$), indicating that older children tend to pronounce more words. There was

also a weak negative correlation between age and fronting (r = -0.19, p = 0.042), suggesting that older children exhibit less fronting. Additionally, we found a weak positive correlation between stopping and fronting (r = 0.18, p = 0.045), indicating that children who use the stopping process more also tend to use fronting more, and vice versa. No other significant correlations were found.

Further analysis by age group revealed specific correlations:

- For the 3-4 year age group, a significant positive correlation was found between the percentage of pronounced words and the use of final consonant deletion (r = 0.40, p = 0.03), as shown in Table C.8 in Appendix C. This indicates that children in this age range who pronounce more words also tend to use the final consonant deletion process more frequently.
- In the 4-5 year age group, Table C.9 in Appendix C shows a significant positive correlation between age and the number of pronounced words (r = 0.37, p = 0.017), indicating that older children in this group tend to pronounce more words. Additionally, there was a positive correlation between stopping and fronting (r = 0.39, p = 0.01), suggesting that children who use stopping more also use fronting more, and vice versa.
- No significant correlations were found in the 5-6 year age group, as presented in Table C.10 in Appendix C.

These analyses are crucial for better understanding how the PPs modeled in this research can distinguish TD children and those with DLD. By examining the correlations between phonological processes and other variables, such as the percentage of repeated words and age, we can identify patterns that may differentiate TD children from those with DLD.

For instance, the significant positive correlation between the percentage of repeated words and the use of final consonant deletion in the 3-4 year age group suggests that children with DLD might exhibit a higher frequency of certain phonological processes. Similarly, the positive correlation between

stopping and fronting in the 4-5 year age group provides insights into the co-occurrence of phonological processes.

By identifying these patterns, we can better understand the phonological characteristics that distinguish TD children from those with DLD. This understanding is essential for developing accurate classification models and diagnostic tools for DLD based on phonological processes.

However, even if these analyses revealed few significant results, mostly because of the small size of our data, our algorithm demonstrated its ability to precisely categorize the PPs used by each child and provide commonly used measures in a format similar to that outputted by KAT.

With the precise description of processes used by children provided by this model, we can evaluate how well the next model, using this output, can classify children as TD or with DLD based on the information provided above.

6.2 MaxEnt results

This section presents the classification results of the MaxEnt model, which identified children as TD or with DLD across different age ranges based on the PPs used, transformed into constraints inputted into the model.

As described in section 5.4.2, an ROC (Receiver Operating Characteristic) curve was plotted for each age range, and the AUC (Area Under the Curve) was calculated to evaluate the model's overall performance. Additionally, the best threshold for classification was determined for each ROC plot.

The threshold in this context refers to the probability value that differentiates between TD and DLD pronunciation patterns. By setting this threshold, we establish the point at which the model decides whether a child's pronunciation pattern is more likely to be from a TD child or from a child with DLD.

In the ROC curve, we plotted the mean and median probabilities of

words pronounced by each child. Since the median is more robust to outliers, we initially compared whether there was a significant difference between the mean and median when plotting the ROC. Finding no significant difference, we concluded that there were no outliers, and thus the mean could be reliably used for our analysis.

Figures 6.4, 6.5, and 6.6 display the ROC curves for each age range. The AUC is a metric used to evaluate the overall performance of the model. It ranges from 0 to 1, where a higher value indicates better model performance in distinguishing between TD children and those with DLD. In our case, the AUC values increased with age, reflecting improved model performance as the children grew older. This trend aligns with expectations, as diagnosing DLD is more challenging at age 3;0, and becomes easier as children develop further.

To determine the optimal classification thresholds, we used the ROC curves. The optimal threshold is the point that achieves the best balance between sensitivity (i.e., the ability to correctly identify true positives) and specificity (i.e., the ability to correctly identify true negatives). We selected the threshold for each age range by finding the point on the ROC curve that maximizes Youden's J statistic (Sensitivity + Specificity - 1) for each age group. This method ensures that we achieve the highest possible accuracy in classifying children. The optimal thresholds were found to be 0.32 for 3-4 year-olds, 0.38 for 4-5 year-olds, and 0.40 for 5-6 year-olds.

In an effort to enhance the accuracy of our classifier, we conducted additional tests. First, we explored the impact of random weight initialization on our model's performance. Previously, the model's weights were initialized to zero, but for these tests, we initialized the weights randomly. We trained the model 10 times with different random initialization for the first group of children (aged 3-4 years) without using k-fold cross-validation. This approach was intended to evaluate the potential benefits of random initialization. The trained models were then tested on a separate test set containing both TD children and children with DLD. We compared the performance of each classifier to assess the impact of random weight initialization.

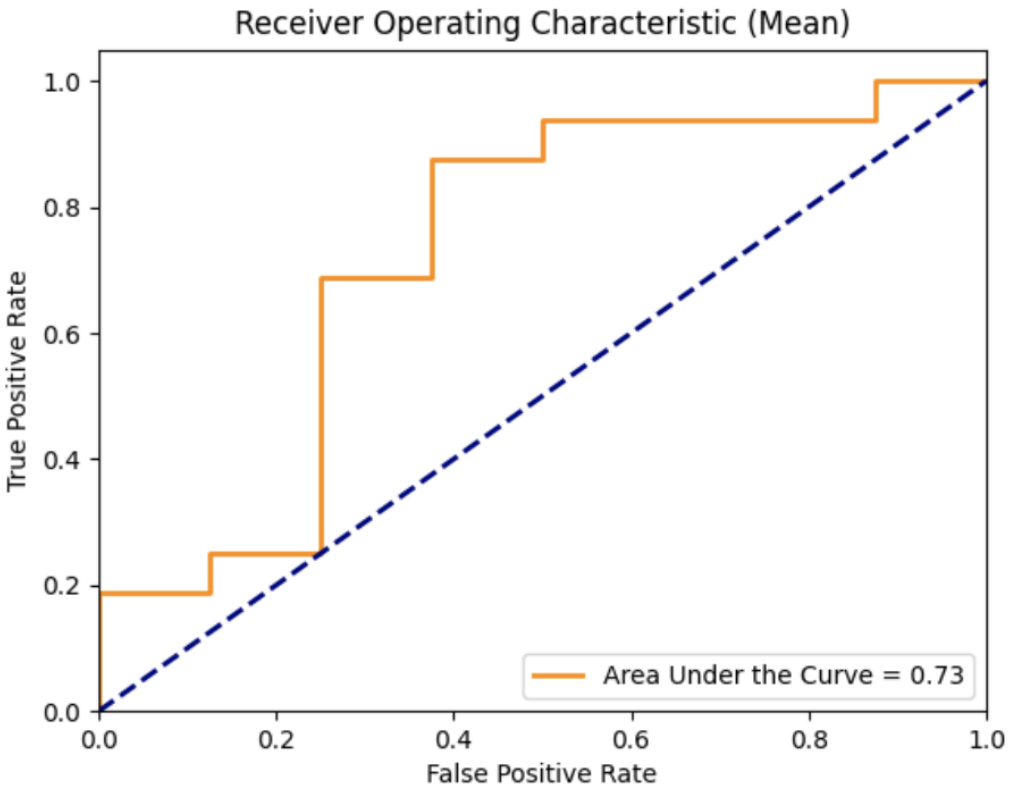

Figure 6.4: ROC plot for the mean probability of children in the test set of the 3 to 4 years old

Testing random weight initialization is interesting because it can help avoid issues related to poor convergence that might occur with weights initialized to zero. Randomly initializing weights can lead to different starting points for the training process, potentially improving the robustness and overall performance of the model.

Next, we examined the effect of varying the L2 regularization parameter. L2 regularization helps to prevent overfitting by penalizing large weights, thereby encouraging the model to learn simpler, more general patterns. The default variance for L2 regularization in our implementation was set to 1000. To determine if a lower variance would improve performance, we ran the model with variances ranging from 1000 down to 100, decrementing by 100 each time, while keeping the weights initialized to zero. The trained weights for each variance were then tested on our test set, and we compared the performance across different variances.

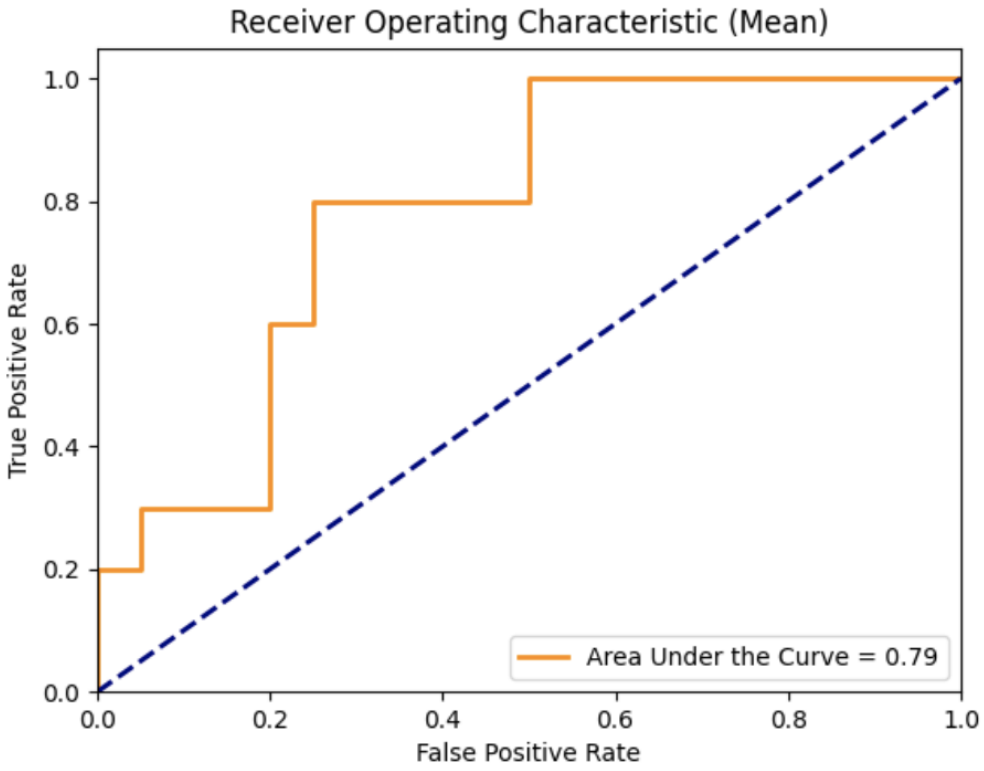

Figure 6.5: ROC plot for the mean probability of children in the test set of the 4 to 5 years old

Testing lower variance values for L2 regularization is interesting because a high variance might be too lenient, allowing the model to overfit the training data. Conversely, a lower variance increases the penalty for large weights, which can help the model generalize better by preventing it from fitting noise in the training data. By systematically decreasing the variance, we aimed to find an optimal balance that minimizes overfitting while maintaining good predictive performance.

However, in both sets of experiments—random weight initialization and varying L2 regularization variances—we found no significant differences in performance compared to our baseline model (with weights initialized to zero and a regularization variance of 1000). Consequently, we concluded that re-training the model with k-fold cross-validation, incorporating these modifications, would not yield any additional benefits, as all tests returned similar performance.

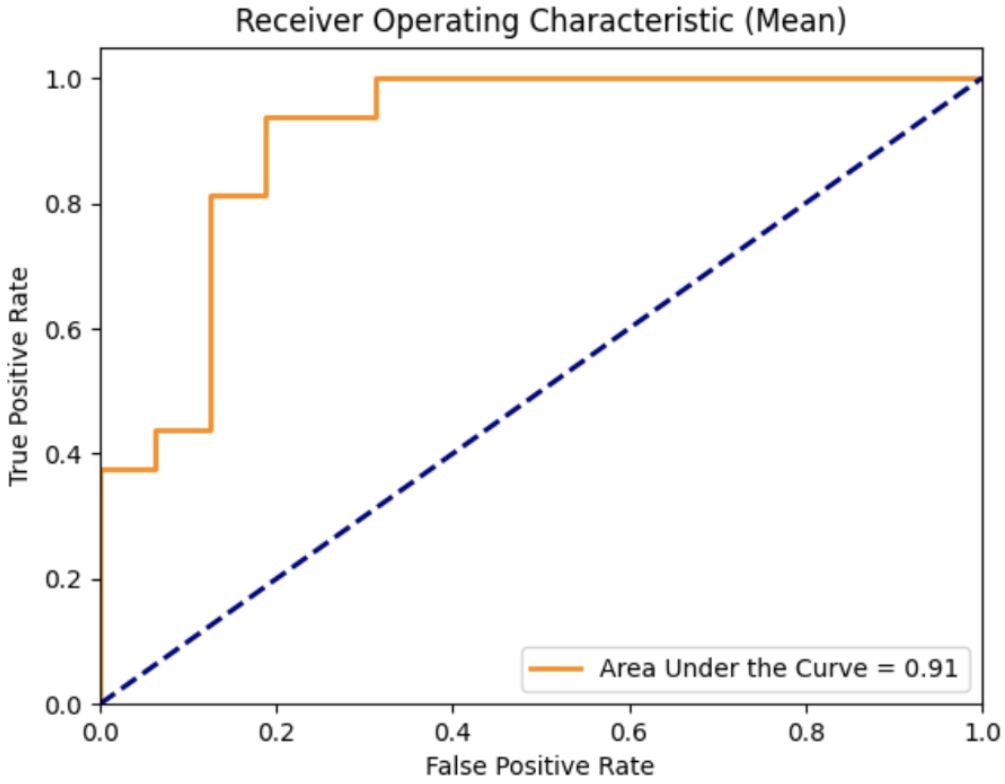

Figure 6.6: ROC plot for the mean probability of children in the test set of the 5 to 6 years old

The fact that initial values and regularization values did not have an impact essentially means that our models do not rely extensively on these values. Moreover, as the results stayed the same, it indicates that our results are robust. The non-existent impact of modifying the initial weights or the variance of the L2 regularization can be explained by the fact that we did not specify any hidden structure, such as syllabic structure, in our MaxEnt models so the optimization space is likely convex with a single optimum. In this study, the input given to these models implemented relatively few constraints without any very complex constraint interactions, so the weighting conditions for the constraints to arrive at the correct pattern should not be so complex that regularization can prevent the algorithm from finding them.

7. Discussion

7.1 Limitations and further work

Data Limitation

One significant limitation of our study was the lack of access to audio data, which prevented the integration of prosodic information into our analyses. The small sample size also limited the power of our statistical analyses, resulting in few significant effects being observed, even though such effects have been demonstrated in previous studies comparing typically developing (TD) children and those with developmental language disorder (DLD).

Classifier Performance

The limited size of our dataset also affected the training of our classifier, which might explain its performance. Despite achieving a reasonable performance with the small dataset and by analyzing only four phonological processes (PPs), the classifier's accuracy could be improved by including more PPs, providing a richer set of information for classification.

Models Implementation

While our models do not require extensive computational resources and can be run on a local laptop, they are not yet user-friendly for speech and language pathologists (SLPs) in their current form due to the lack of a designed interface. These models are not integrated with existing diagnostic tools, which is an area for future work. However, they hold potential to aid in clinical decision-making once these usability issues are addressed.

Generalizability and Applicability

Our models can be generalized to any phonetic transcription with reference and prediction words, allowing for direct phonetic analysis rather than relying solely on pseudo-phonetic transcriptions. Even if, for the purpose of this research, an algorithm has been developed to convert pseudo-phonetic transcriptions into real phonetic ones as a pre-processing step if needed. However, the current approach is limited to matching word pairs and does not accommodate spontaneous speech, which should be explored in future work.

Additionally, while our models were tailored for Dutch, they could potentially be adapted to other languages, except for the phonetic conversion component. The architecture of the models is supposed to make them robust to variations in input data quality, such as differences in pronunciation accuracy or dialectal variations, but this needs further investigation and training with more diverse data to confirm.

Interpretability

The LD-BFS and the algorithm outputting the PPs are designed to be highly interpretable, making their outputs accessible to other researchers and clinicians.

The MaxEnt model, however, requires plotting on a ROC curve to determine an optimal threshold. Users must then calculate the mean probability of all words pronounced by a child to obtain a result, which is not straightforward. Nonetheless, with proper explanation, the outputs can be interpreted easily.

Futures Directions

Future work should focus on several key areas to enhance the robustness and applicability of our models.

Integrating prosodic information by gaining access to audio data would allow for a more comprehensive analysis, capturing features beyond seg-

mental phonology. Using, for example, tools such as PhonChild [43] or AASP [44], allowing the analysis of phonetic transcriptions along with speech audio to augment the transcription with more comprehensive prosodic annotations.

Increasing the sample size is critical, as larger datasets would improve the power of our statistical analyses and the performance of our classifier.

Additionally, enhancing the usability of the models is essential. Collaborating with professionals who specialize in developing user-friendly interfaces and integrating these models with existing diagnostic tools could make them more accessible and practical for SLPs. Such interdisciplinary collaboration would ensure that the models developed in this research can be effectively implemented in clinical settings.

Expanding the analysis to include more PPs could provide a richer set of information, improving the classifier's accuracy.

Adapting the models to handle spontaneous speech, rather than just matching words, could offer more naturalistic insights into children's language abilities.

Moreover, cross-linguistic adaptation of the models would broaden their applicability, allowing for use with different languages and dialects.

Moreover, another direction would be to develop a longitudinal model by exploring how the frequency of examples encountered influences the ranking of the Optimality Theory constraints. This exploration could model a gradual transition from one stage of language acquisition to another.

Finally, further research is needed to evaluate the models' robustness to variations in input data quality, such as differences in pronunciation accuracy or dialectal variations, ensuring they can perform reliably across diverse scenarios.

Therefore, while our models show promise, particularly in terms of interpretability and low computational demands, there is significant scope for enhancing their practical application and performance through further de-

velopment and validation.

7.2 Conclusion

This research aimed to use a combined analysis of phonological processes and classification outcomes to support speech and language pathologists in their diagnostic and decision-making process for tailored therapy regarding children's language development and impairment.

We used a non-word repetition task dataset provided by Auris from Dutch typically developing children and children with developmental language disorder between age 3;0 and 6;3.

As no current automatic method is able to phonetically transcribe (Dutch) child speech, we used the existing pseudo-phonetic transcription given by Auris and automatically translated it into real phonetic transcription. If this pseudo-phonetic transcription has limits when transcribing spontaneous speech, as the non word repetition task uses pseudowords designed by researchers, with each grapheme representing one exact phoneme, we were able to automatically transcribe the pseudo-phonetic into real phonetic as input to our models. This allowed us to prepare our models for future integration with automatic speech recognition systems.

In this research, we first aimed to accurately model phonological processes from phonetic transcriptions of Dutch children's speech. By employing Levenshtein distance and Breadth-First Search algorithms, we successfully identified four common phonological processes: final consonant deletion, stopping, fronting, and gliding. This modeling provided detailed and interpretable analyses for individual children and age groups.

Secondly, we investigated whether the modeled phonological processes could accurately distinguish between typically developing children and those with developmental language disorder (DLD) in our data. The statistical analysis of the phonological processes alone did not reveal significant

differences between these groups, likely due to the limited dataset and the narrow scope of analyzed processes.

Thirdly, we assessed the effectiveness of a classifier trained on the modeled phonological processes in distinguishing between typical children and those with DLD. Our Maximum Entropy classifier demonstrated promising accuracy, with performance ranging from 73% to 91%. This suggests that such a classifier can be an effective tool for distinguishing between typical children and those with DLD, especially when refined and supported by larger datasets.

Fourthly, we examined the reliability of the modeled phonological processes and classifiers across different age groups of Dutch children. Our findings indicated that the accuracy of classification improves with age, highlighting the potential for more reliable diagnostics in older children.

Despite the challenges and limitations, this research provides a solid foundation for future work in automated phonological process analysis and classification in child speech, aiming to support tailored therapeutic interventions for children with language impairments.

Appendices

A. Appendix A

	Process	Realisation	Target	Gloss
a)	Cluster Reduction	te	twe	two
	Cl → CØ	klm@	klIm@	climb
	sC → ØC	sOm@	sxOm@l	swing
	Cr → CØ	xOn	xrOnt	ground
	CCC → C(C) # I:s	# Irst	# first	
	Final C Deletion	tO	tOt	until
	Reduplication	bumbum	blum	flower
	Pretonic Syllable Deletion	nej@	b@'ned@	down
	Regressive Assimilation	tIt@	sIt@	to sit
	Devoicing	pʌ	bʌI	ball
	Fronting of Velars	hOn@r	hOŋ@r	hunger
	Gliding	wam@	ram@	windows
	Stopping	pat	blas@	to blow
b)	Cluster Deletion	ka:	kʌnt	side
	Cluster Creation	honde	x@won@	normal
	Initial C Deletion	ɔül@	wil@	wheels
	Metathesis	plIsti	politsi	police
	Posttonic Syllable Deletion	xo	xoi@	throw
	Progressive Assimilation	lʌl@	ʌŋ@	long
	Depalatalisation	sirʌf	širʌf	giraf
	Denasalisation	bIa	mIr	more
	Delabialisation	sInt	fInt	find
	Frication	fuk	buk	book
	Glottalisation	bE?	bEn	am
	H-zation	hut@	mut@	must
	Labial Lenition	wat	part	horse
	Nasalisation	neç@	des@	this
	Palatalisation	šwʌtˢ	swʌrt	black
	Velarisation	xʌnt	sʌnt	sand
	Vocalisation	hoxo	fox@l	bird
	Voicing	bus	pus	cat
c)	Initial C Addition	k@'len	ʌ'len	alone
	Lateralisation	slEm@	swEm@	to swim

Figure A.1: Processes applied by the Dutch phonologically impaired children, with examples from Beers [17] data

B. Appendix B

1 syllable words	2 syllable words	3 syllable words	4 syllable words	5 syllable words
Jaat	Hoolin	Sietaalon	Moolekaatus	Wookaaloemoodon
Luup	Hiemup	Luujeemuk	Wuiseujoenif	Soegonuifeusir
Jiek	Keupun	Poekuijol	Kuuwoebeujeg	Nujigeufuusut
Peek	Naatep	Peelaanot	Kootaabielan	Beemonievoekes
Peun	Nuipok	Suitaajin	Meufienoegir	Jeunimeusuifir
Loen	Keepon	Liepoetaan	Siewaatoolan	Geeruutievaanot

Figure B.1: Target Words

C. Appendix C

Shapiro-Wilk Test

Descriptives

	diagnosis	Percentage Final C Del	Percentage Pronounced words	Percentage Fronting	Percentage Stopping	Percentage Gliding
N	DLD	50	50	50	50	50
	TD	62	62	62	62	62
Shapiro-Wilk W	DLD	0.348	0.963	0.482	0.272	0.536
	TD	0.520	0.887	0.206	0.176	0.608
Shapiro-Wilk p	DLD	<.001	0.114	<.001	<.001	<.001
	TD	<.001	<.001	<.001	<.001	<.001

Figure C.1: Shapiro Wilk Test

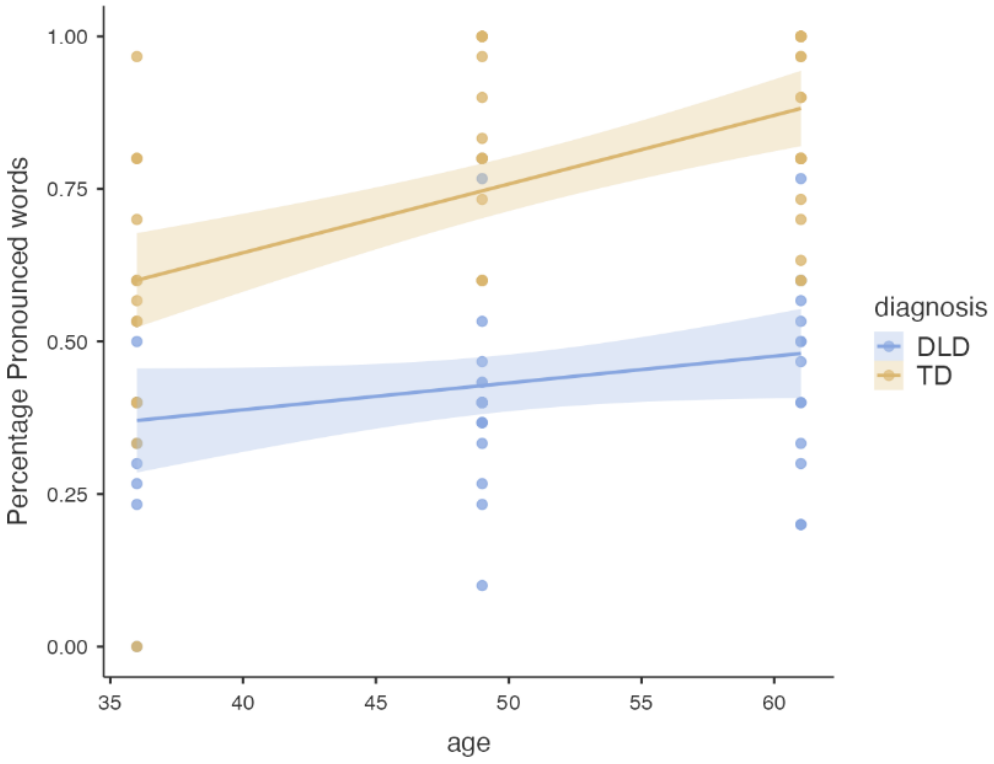

Figure C.2: Visualisation of the use of the percentage of pronounced words across age and groups

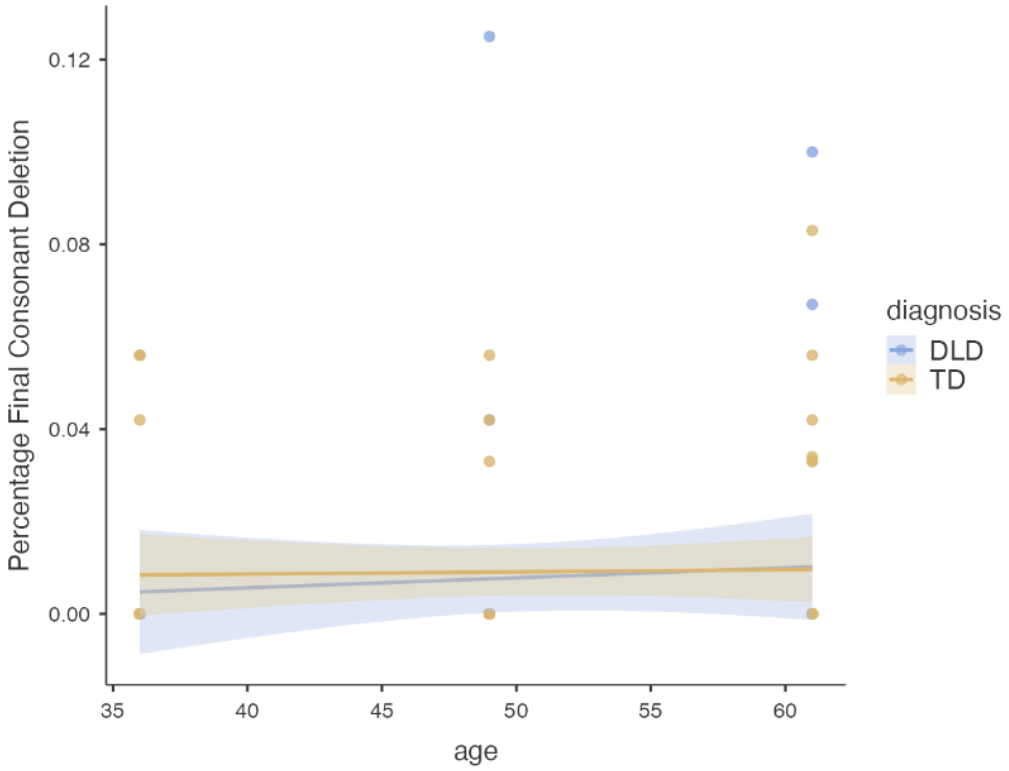

Figure C.3: Visualisation of the percentage of use of the final consonant deletion process across ages and groups

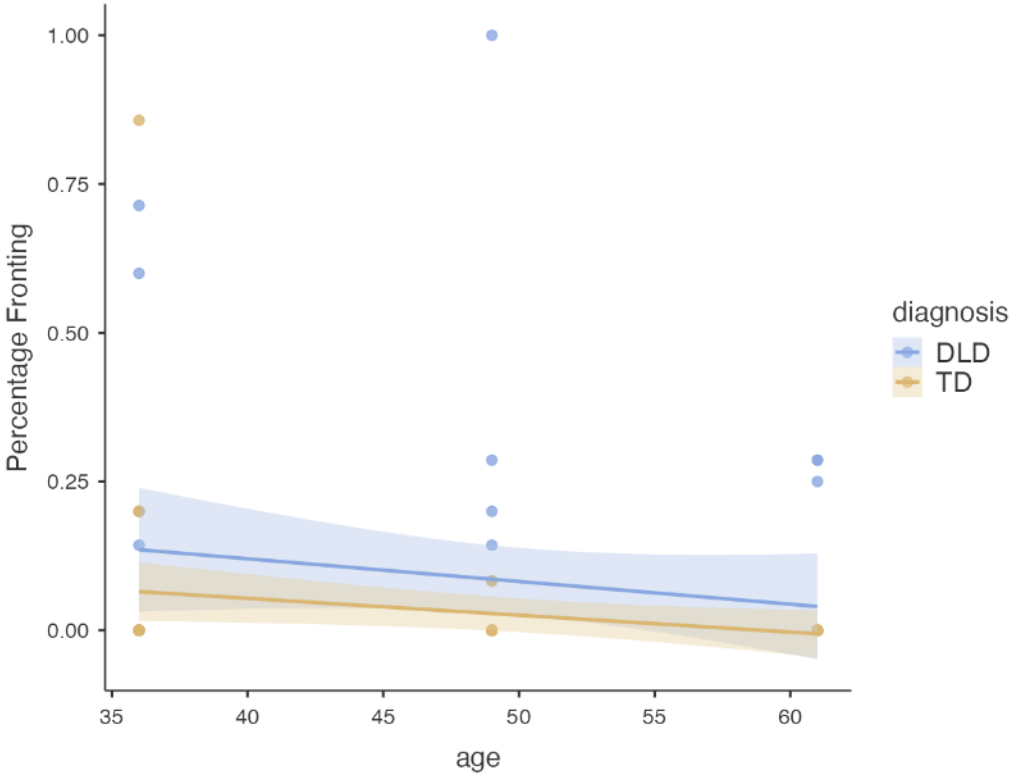

Figure C.4: Visualisation of the percentage of use of the fronting process across ages and groups

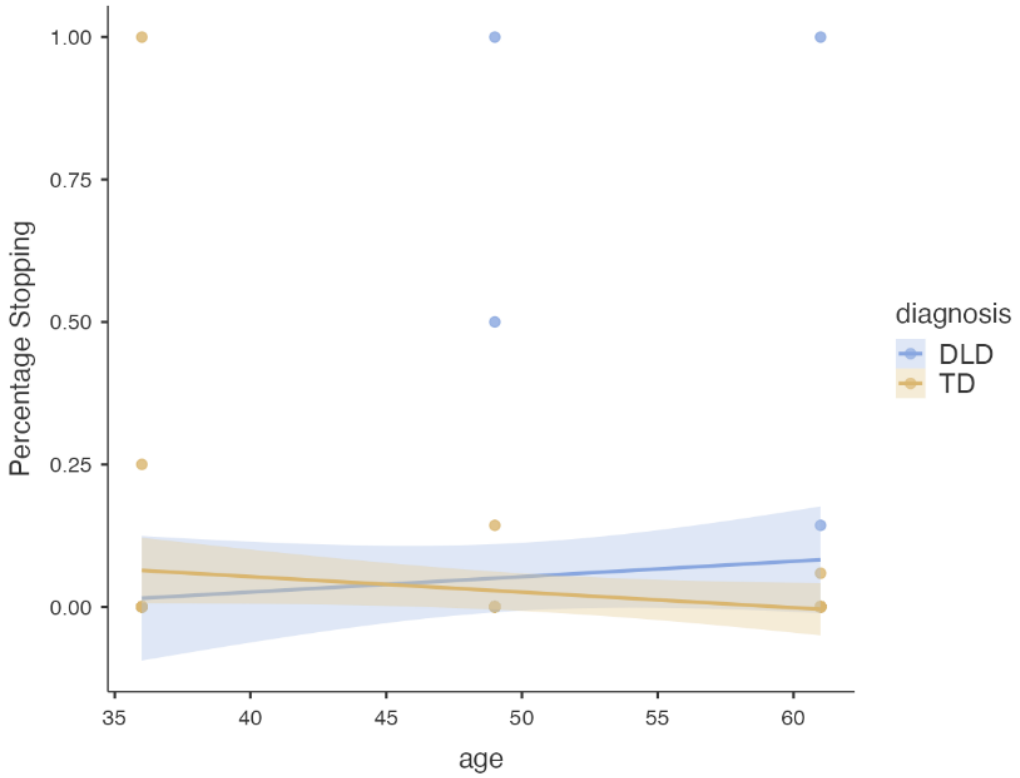

Figure C.5: Visualisation of the percentage of use of the stopping process across ages and groups

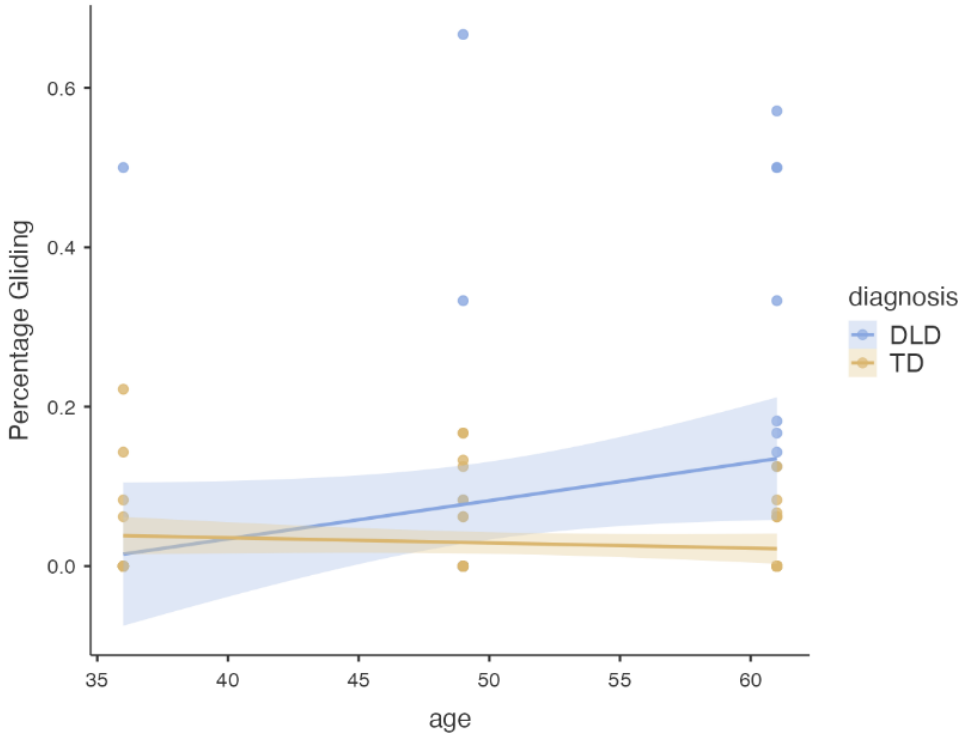

Figure C.6: Visualisation of the percentage of use of the gliding process across ages and groups

Correlation Matrix

		Percentage Pronounced words	Percentage Final Consonant Deletion	Percentage Fronting	Percentage Stopping	Percentage Gliding	age
Percentage Pronounced words	Spearman's rho	—					
	df	—					
	p-value	—					
Percentage Final Consonant Deletion	Spearman's rho	0.129	—				
	df	110	—				
	p-value	0.175	—				
Percentage Fronting	Spearman's rho	−0.175	0.036	—			
	df	110	110	—			
	p-value	0.065	0.708	—			
Percentage Stopping	Spearman's rho	0.024	0.155	0.189 *	—		
	df	110	110	110	—		
	p-value	0.800	0.103	0.045	—		
Percentage Gliding	Spearman's rho	0.076	0.099	−0.055	0.096	—	
	df	110	110	110	110	—	
	p-value	0.425	0.298	0.565	0.314	—	
age	Spearman's rho	0.362 ***	0.037	−0.193 *	−0.010	0.120	—
	df	110	110	110	110	110	—
	p-value	<.001	0.699	0.042	0.917	0.209	—

Note. * p < .05, ** p < .01, *** p < .001

Figure C.7: Spearsman Correlation Matrix for children across all ages presenting the correlation between the different percentage of phonological processes use and age

Correlation Matrix

		Percentage Pronounced words	Percentage Final Consonant Deletion	Percentage Fronting	Percentage Stopping	Percentage Gliding
Percentage Pronounced words	Spearman's rho	—				
	df	—				
	p-value	—				
Percentage Final Consonant Deletion	Spearman's rho	0.407 *	—			
	df	26	—			
	p-value	0.032	—			
Percentage Fronting	Spearman's rho	−0.029	0.249	—		
	df	26	26	—		
	p-value	0.883	0.202	—		
Percentage Stopping	Spearman's rho	0.159	0.311	0.239	—	
	df	26	26	26	—	
	p-value	0.420	0.107	0.220	—	
Percentage Gliding	Spearman's rho	0.246	−0.189	−0.264	−0.128	—
	df	26	26	26	26	—
	p-value	0.206	0.336	0.174	0.515	—

Note. * p < .05, ** p < .01, *** p < .001

Figure C.8: Spearsman Correlation Matrix for 3-4 years old children presenting the correlation between the different percentage of phonological processes use and age

Correlation Matrix

		Percentage Pronounced words	Percentage Final Consonant Deletion	Percentage Fronting	Percentage Stopping	Percentage Gliding	age
Percentage Pronounced words	Spearman's rho	—					
	df	—					
	p-value	—					
Percentage Final Consonant Deletion	Spearman's rho	0.103	—				
	df	38	—				
	p-value	0.528	—				
Percentage Fronting	Spearman's rho	−0.020	0.027	—			
	df	38	38	—			
	p-value	0.902	0.870	—			
Percentage Stopping	Spearman's rho	0.037	0.158	0.391 *	—		
	df	38	38	38	—		
	p-value	0.822	0.331	0.013	—		
Percentage Gliding	Spearman's rho	0.246	0.269	0.286	0.064	—	
	df	38	38	38	38	—	
	p-value	0.125	0.093	0.073	0.696	—	
age	Spearman's rho	0.376 *	0.086	0.048	0.116	−0.063	—
	df	38	38	38	38	38	—
	p-value	0.017	0.597	0.768	0.474	0.698	—

Note. * p < .05, ** p < .01, *** p < .001

Figure C.9: Spearsman Correlation Matrix for 4-5 years old children presenting the correlation between the different percentage of phonological processes use and age

Correlation Matrix

		Percentage Pronounced words	Percentage Final Consonant Deletion	Percentage Fronting	Percentage Stopping	Percentage Gliding	age
Percentage Pronounced words	Spearman's rho	—					
	df	—					
	p-value	—					
Percentage Final Consonant Deletion	Spearman's rho	0.024	—				
	df	42	—				
	p-value	0.876	—				
Percentage Fronting	Spearman's rho	−0.251	−0.127	—			
	df	42	42	—			
	p-value	0.100	0.413	—			
Percentage Stopping	Spearman's rho	0.004	0.065	−0.073	—		
	df	42	42	42	—		
	p-value	0.979	0.676	0.637	—		
Percentage Gliding	Spearman's rho	−0.294	0.116	−0.180	0.259	—	
	df	42	42	42	42	—	
	p-value	0.053	0.453	0.241	0.090	—	
age	Spearman's rho	−0.211	−0.022	−0.105	−0.028	0.018	—
	df	42	42	42	42	42	—
	p-value	0.169	0.888	0.499	0.858	0.906	—

Note. * p < .05, ** p < .01, *** p < .001

Figure C.10: Spearsman Correlation Matrix for 5-6 years old children presenting the correlation between the different percentage of phonological processes use and age

Bibliography

[1] P. Fikkert, "Acquisition of phonology," in *The first Glot International state-of-the-article book. The latest in linguistics*, ser. Studies in Generative Grammar, L. Cheng and R. Sybesma, Eds., vol. 48, Berlin/New York: Mouton de Gruyter, 2000, pp. 221–250.

[2] J. Verhagen, J. Boom, H. Mulder, E. de Bree, and P. Leseman, "Reciprocal relationships between nonword repetition and vocabulary during the preschool years," *Developmental Psychology*, vol. 55, no. 6, p. 1125, 2019.

[3] A. Prince and P. Smolensky, "Optimality theory: Constraint interaction in generative grammar," *Optimality Theory in phonology: A reader*, pp. 1–71, 2004.

[4] P. Fikkert, "The acquisition of dutch phonology," *PRAGMATICS AND BEYOND NEW SERIES*, pp. 163–222, 1998.

[5] M. M. Vihman, *Phonological development: The origins of language in the child*. Blackwell Publishing, 1996.

[6] D. Ingram, "Phonological rules in young children," *Journal of child language*, vol. 1, no. 1, pp. 49–64, 1974.

[7] N. V. Smith, *The acquisition of phonology: A case study*. Cambridge University Press, 1973.

[8] D. Ingram, "Phonological disability in children," 1976.

[9] J. V. Irwin and S. P. Wong, "Phonological development in children 18 to 72 months," *(No Title)*, 1983.

[10] J. L. Locke and M. Studdert-Kennedy, *Phonological acquisition and change*. Academic Press New York, 1983.

[11] P. Grunwell, "The development of phonology: A descriptive profile," *First language*, vol. 2, no. 6, pp. 161–191, 1981.

[12] A. S. S. Gillis, *The language acquisition of the child, a renewed development in dutch-language research*, 1987.

[13] W. Beers, "The phonology of normally developing and language-impaired children," 1995.

[14] I. Mennen, C. Levelt, and E. Gerrits, "Acquisition of dutch phonology: An overview," *QMU Speech Science Research Centre Working Papers*, 2006.

[15] L. van Haaften, S. Diepeveen, L. van den Engel-Hoek, B. de Swart, and B. Maassen, "Speech sound development in typically developing 2–7-year-old dutch-speaking children: A normative cross-sectional study," *International journal of language & communication disorders*, vol. 55, no. 6, pp. 971–987, 2020.

[16] C. R. Marshall, "Word production errors in children with developmental language impairments," *Philosophical Transactions of the*

Royal Society B: Biological Sciences, vol. 369, no. 1634, p. 20 120 389, 2014.

[17] M. Beers, "Phonological processes in dutch language impaired children," *Scandinavian journal of logopedics and phoniatrics*, vol. 17, no. 1, pp. 9–16, 1992.

[18] N. Chomsky and M. Halle, *The Sound Pattern of English*. New York: Harper & Row, 1968.

[19] R. Kager, J. Pater, and W. Zonneveld, *Constraints in phonological acquisition*. Cambridge University Press, 2004.

[20] J. A. Barlow, "Case study: Optimality theory and the assessment and treatment of phonological disorders," *Language Speech and Hearing Services in Schools*, vol. 32, no. 4, pp. 242–256, 2001.

[21] D. Archangeli, "Optimality theory: An introduction to linguistics in the 1990s," *Optimality theory: An overview*, pp. 1–32, 1997.

[22] C. A. Ferguson and O. K. Garnica, "Theories of phonological development," in *Foundations of language development*, Elsevier, 1975, pp. 153–180.

[23] J. A. Barlow and J. A. Gierut, "Optimality theory in phonological acquisition," *Journal of Speech, Language, and Hearing Research*, vol. 42, no. 6, pp. 1482–1498, 1999.

[24] R. Kager, J. Pater, and W. Zonneveld, *Constraints in phonological acquisition*. Cambridge University Press, 2004.

[25] F. Shooshtaryzadeh, "Optimality theory and assessment of developing and disordered phonologies," *Journal of Indian Speech Language & Hearing Association*, vol. 29, no. 2, pp. 13–20, 2015.

[26] J. A. Gierut, "Treatment efficacy: Functional phonological disorders in children," *Journal of Speech, Language, and Hearing Research*, vol. 41, no. 1, S85–S100, 1998.

[27] J. A. Gierut, "Complexity in phonological treatment," 2001.

[28] G. Jones, F. Gobet, and J. M. Pine, "Linking working memory and long-term memory: A computational model of the learning of new words," *Developmental Science*, vol. 10, no. 6, pp. 853–873, 2007.

[29] M. Tamburelli, G. Jones, F. Gobet, and J. M. Pine, "Computational modelling of phonological acquisition: Simulating error patterns in nonword repetition tasks," *Language and Cognitive Processes*, vol. 27, no. 6, pp. 901–946, 2012.

[30] M. Shahin, U. Zafar, and B. Ahmed, "The automatic detection of speech disorders in children: Challenges, opportunities, and preliminary results," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 400–412, 2019.

[31] H. T. Bunnell, D. Yarrington, and J. B. Polikoff, "Star: Articulation training for young children.," in *INTERSPEECH*, Citeseer, 2000, pp. 85–88.

[32] M. H. Franciscatto, M. D. Del Fabro, J. C. D. Lima, *et al.*, "Towards a speech therapy support system based on phonological processes

early detection," *Computer speech & language*, vol. 65, p. 101 130, 2021.

[33] L. Ward, A. Stefani, D. V. Smith, *et al.*, "Automated screening of speech development issues in children by identifying phonological error patterns.," in *Interspeech*, 2016, pp. 2661–2665.

[34] A. Ratnaparkhi, "A maximum entropy model for part-of-speech tagging," in *Conference on empirical methods in natural language processing*, 1996.

[35] V. Sahayak, V. Shete, and A. Pathan, "Sentiment analysis on twitter data," *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*, vol. 2, no. 1, pp. 178–183, 2015.

[36] B. Hayes and C. Wilson, "A maximum entropy model of phonotactics and phonotactic learning," *Linguistic inquiry*, vol. 39, no. 3, pp. 379–440, 2008.

[37] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 939–952, 1982.

[38] S. Schneider, A. Baevski, R. Collobert, and M. Auli, "Wav2vec: Unsupervised pre-training for speech recognition," *arXiv preprint arXiv:1904.05862*, 2019.

[39] A. van der Klis, F. Adriaans, M. Han, and R. Kager, "Using open-source automatic speech recognition tools for the annotation of dutch infant-directed speech," *Multimodal Technologies and Interaction*, vol. 7, no. 7, p. 68, 2023.

[40] P. M. Skaer, "Universal grammar, optimality theory and first language acquisition," *Memoirs of the Faculty of Integrated Arts and Sciences, Hiroshima University. V, Studies in Linguistic Culture*, vol. 29, pp. 19–44, 2003.

[41] D. Ingram, "The measurement of whole-word productions," *Journal of Child Language*, vol. 29, no. 4, pp. 713–733, 2002.

[42] E. Babatsouli, D. Ingram, and D. Sotiropoulos, "Phonological word proximity in child speech development," *Chaotic Modeling and Simulation*, vol. 4, no. 3, pp. 295–313, 2014.

[43] Y. Rose, "Childphon: A database solution for the study of child phonology," in *Proceedings of the 27th annual Boston University conference on language development*, 2003, pp. 674–685.

[44] N. Hu, B. Janssen, J. Hansen, C. Gussenhoven, and A. Chen, "Automatic analysis of speech prosody in dutch," 2020.