The Identification of Typical and Atypical Phonological Acquisition in Turkish-German Bilingual Children

Katharina Margareta Albrecht

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield
Faculty of Medicine, Dentistry and Health
Department of Human Communication Sciences

August 2017

ABSTRACT

Bilingual phonological acquisition is special as it encompasses the parallel processing of more than one phonological system and exhibits an interaction between the involved languages. Its assessment is mainly impeded by the heterogeneity of the bilingual population and the unavailability of suitable assessments; hence knowledge of what constitutes a typical bilingual phonological acquisition is generally limited for most language combinations. This knowledge, however, would enhance the understanding of phonological development in general and form the basis for the identification of atypical development. Since no study has yet comprehensively investigated the typical phonological acquisition in Turkish-German bilingual children and differentiated it from atypical acquisition, the present thesis aimed to explore these aspects by assessing the phonological skills of 84 Turkish-German bilingual children aged 3;0 - 5;5 years in both languages (t1) and following up 43 of these participants 12 - 15 months after the initial assessment (t2). Additionally, performances on 'quasi-language-independent' psycholinguistic tasks were assessed at t2 and evaluated regarding their significance for differentiating typical from atypical development.

Analyses revealed that the typical phonological acquisition in Turkish-German bilinguals included an overall slower rate and qualitative differences compared to that in monolingual children, dissimilar phonological skills in German and Turkish (also over time), an interaction between the two languages, an influence by child-internal and environmental factors as well as a general improvement over time. Regarding the identification of potential quantitative and qualitative markers for the differentiation of typical from atypical development, three factors could be determined: the nature of children's phonological patterns, their number of infrequent variants as well as their performances on psycholinguistic tasks. Further evaluations showed, however, that only the combination of these markers as well as a longitudinal monitoring of children's performances allowed for a reliable differentiation within the present cohort. Theoretical and clinical implications of these outcomes are discussed.

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my PhD supervisors, Dr Silke Fricke, Prof. Annette Fox-Boyer, Dr Jane McCormack and Prof. Joy Stackhouse. The faith they had in me and the personal and academic support they provided me with throughout the past years is truly exceptional and deserves my greatest respect and thankfulness. I have learned a lot from their feedback as well as our thought-provoking discussions and can count myself lucky that I had all of them as my mentors.

Many thanks also go to the cooperative nurseries, interested parents, and lovely children who kindly participated in this research and helped to make this project possible.

Special thanks are due to my test assistants Tuğba Gürleyik, Deniz Okyay, Leyla Balkas, Elif Uyulur, Başak Uzun und Büşra Yildiz without whom none of the assessments in Turkish could have taken place. Additionally, I would like to thank Sevil Czesla, Merva Tan and Ayşe Kelcikoğlu who did a fantastic job in translating the information sheets, consent forms and parental questionnaires into Turkish.

I am very thankful for the time and effort Sigrun Lang, Laura Kubaschk and Irene Trapp have put into transcribing and analysing my data for the inter-rater reliability analyses.

Also, I would like to thank my dear fellow PhD colleagues who have lent me an ear when required, offered support wherever they could and significantly contributed to making my PhD life a very pleasant experience. Especially, I would like to thank Dr Marit Clausen and Sigrun Lang for the endless discussions on phonological theories, phonetic transcription and analysis-related issues, which were so unbelievably helpful.

Further, I am very thankful for all the wonderful and generous people who have offered me a place to stay during PhD-times I spent in the UK. Helen Cameron, Silke Fricke, Rebekka Niepelt & Sotirios Karampampas, Alexandra & Paul Oseghale as well as my “English family” have always made me feel very much at home.

Special thanks also go to my friends Katy Syme and Helen Cameron for proofreading.

Moreover, I would also like to express my gratitude to my parents and brothers who have supported me in so many different ways during the last years, listened to my dreams, hopes, wishes and worries, and never stopped encouraging and believing in me.

Special thanks also go to my dear friend Alina Drooff who was never short on encouraging words, sympathy and love throughout this time.

Finally, much love and deep gratitude go to my wonderful husband, Sebastian, for being the solid rock during ups and downs of this PhD journey, his patience, the endless Excel and Word gimmicks he taught me, his encouraging words, love and never-ending belief in me and what I am doing!

PUBLICATIONS ARISING FROM THIS THESIS

Albrecht, K. M., Fricke, S., Fox-Boyer, A., McCormack, J., & Stackhouse, J. (accepted). *Are Turkish-German bilingual children with weak phonological skills in one language at risk for speech sound difficulties?* Paper at the ISMBS 2017, Chania.

Albrecht, K. M., Fox-Boyer, A., Fricke, S., & Stackhouse, J. (2016, November). *Einsatz von psycholinguistischen Testaufgaben in der Aussprachediagnostik von bilingualen Kindern* [Application of psycholinguistic assessment tasks in the diagnostics of speech sounds in bilingual children]. Paper presented at the ISES IX, Heidelberg.

Albrecht, K. M., Fricke, S., Fox-Boyer, A. & Stackhouse, J. (2016, June). *How informative are psycholinguistic tasks for the identification of speech sound difficulties in Turkish-German bilingual children?* Paper presented at the 16th ICPLA Conference, Halifax.

Albrecht, K. M., Fricke, S., Fox-Boyer, A. & Stackhouse, J. (2016, June). *Development of a phonology assessment for Turkish-German bilingual children living in Germany.* Paper presented at the 16th ICPLA Conference, Halifax.

Salgert, K. M., Fricke, S., Fox-Boyer, A. & Stackhouse, J. (2016, May). *Langzeitdaten zur Ausspracheentwicklung türkisch-deutsch bilingualer Kinder – Welche Informationen liefern sie für die klinische Entscheidungsfindung im Diagnostikprozess?* [Longitudinal data on the speech sound development in Turkish-German bilingual children – What information do they provide for the clinical decision making process?]. Paper presented at the 45th dbl-Jahreskongress, Bielefeld.

Salgert, K., Fricke, S., Stackhouse, J., & Fox-Boyer, A. (2015, September). *Production patterns of consonant clusters in the process of speech acquisition in Turkish-German bilingual children*. Paper at the ISMBS, Chania.

Salgert, K., Fricke, S., Stackhouse, J., & Fox-Boyer, A. (2015, July). *The acquisition of consonants in Turkish-German bilingual children*. Paper presented at the Child Language Symposium, Warwick.

Salgert, K. M., Fricke, S., Stackhouse, J., & Fox-Boyer, A. V. (2014, November). *Erste normative Daten zum Phonologieerwerb türkisch-deutsch bilingualer Kinder* [First normative data on the phonological acquisition in Turkish-German bilingual children]. Paper presented at the ISES VIII, München.

Salgert, K. M., Fox-Boyer, A. V., Fricke, S., & Stackhouse, J. (2014, June). *Erwerb und Störungen der Aussprache bei türkisch-deutsch bilingualen Kindern* [Acquisition and disorders of speech sounds in Turkish-German bilingual children]. Paper presented at the dbl-Jahreskongress, Berlin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AoA	Age of acquisition of a language
DEAP	Diagnostic Evaluation of Articulation and Phonology
CLT	Cross-linguistic transfer
CC(s)	Consonant cluster(s)
InfVar	Infrequent variants
L1 / L2	First language / second language
MLU_w	Mean length of utterance - words
MSAP	Madison Speech Assessment Protocol
NWR	Non-word repetition
PCC	Percentage of consonants correct
PCC-A	Percentage of consonants correct - adjusted
PhonVar	Phonological variations
PLAKSS-II	Psycholinguistic analysis of children’s speech disorders - II
SCREEMIK 2	Screening der Erstsprachkompetenzen bei Migrantenkindern: Russisch-Deutsch, Türkisch-Deutsch 2
SDCS	Speech Disorders Classification System
SLI	Speech and language impairment
SLT	Speech and language therapy/therapist(s)
SSD	Speech sound difficulty/ies
TAT	Türkisch-Artikulations-Test
t1 / t2	time 1 / time 2
TPA	Turkish Phonology Assessment
WIELAU-T	Wiener Lautprüfverfahren für Türkisch sprechende Kinder

CONVENTIONS

Ofen	orthographic realisation of a German or a Turkish word
[ˈʔoːfən]	phonetic realisation of a word, actual realisation of the child
/oːfən/	target phonemic realisation of a word
(oven)	translation of a German or Turkish word into English

Display of age: 3;0 years = 3 years and 0 months

INTRODUCTION

Due to a world-wide increase in migration, especially during the last two decades (United Nations, 2016), the number of people using more than one language in everyday communications has risen remarkably (e.g., Crystal, 2010). Germany is no exception to this trend with 21% of the population (and 35.9% of the children aged 0 - 5 years) having a migration background (Statistisches Bundesamt, 2017). The increasing number of bilingual[1] individuals influences the work of every profession, but especially those working with bilingual children (e.g., nursery nurses, teachers, speech and language therapists [SLT]). Their job is to evaluate children’s communicative and intellectual development which will have a great impact on their future educational attainment, job prospects, emotional well-being and social life (e.g., Lewis, Freebairn, & Taylor, 2000; McCormack, McLeod, McAllister, & Harrison, 2009; Nathan, Stackhouse, Goulandris, & Snowling, 2004b).

Individuals who are bilingual usually use one or more languages at home that differ from the environmental language and are likely to have increased contact to non-native speakers of their languages (Grosjean, 2013). Consequently, their language exposure (input) and usage (output) patterns of each language are quantitatively and qualitatively dissimilar to monolinguals and often involve a later age of acquisition (AoA) of at least one of their languages (Grosjean, 2013; MacWhinney, 2005b; McLeod, Verdon, & IEPMCS, 2017). Thus, bilingual individuals may not be as proficient in all of their languages as their monolingual peers (Goldstein & McLeod, 2012; Grosjean, 1989) and their speech (and language) are likely to exhibit special characteristics of bilingual development (e.g., transfer of linguistic features from one language to the other; Paradis & Genesee, 1996). The potentially weaker language abilities in one or all of their languages as well as the special features of bilingual speech, however, might impede an examiner’s judgment as diversity is not always easily discriminated from difficulty (Kohnert, 2013). A particular challenge exists for paediatricians and SLTs whose task it is to identify typical and atypical speech (and language) development in these children (De Lamo White & Jin, 2011; Gagarina, 2014b). Reliable identification of speech and language difficulties requires comprehensive knowledge about the characteristics of typical and atypical development in a given population, the availability of appropriate (i.e., valid and reliable) assessment tools as well as normative data of typically developing children for comparison (Dodd, 1995; Ingram, 1989b). Despite all of these requirements being fulfilled for monolingual populations of a large number of languages, this is not the

[1] The term *bilingual* is used to refer to people being exposed to and using two (or more) languages in their everyday lives in this research.

case for (most) bilingual populations (De Lamo White & Jin, 2011; Skahan, Watson, & Lof, 2007). The heterogeneity of bilingual speakers makes meeting these requirements particularly challenging as professionals also need to account for the potential influence of the child's bilingual situation (e.g., language input and output patterns), length of exposure to the environmental language as well as children's overall language proficiency in their languages (McLeod, Verdon, & Bowen, 2013).

About 2.85 million people with a Turkish migration background currently live in Germany (Statistisches Bundesamt, 2017). They form the largest subgroup of people with a migration background (i.e., 16.7%) and constitute about 3.5% of the entire population (Statistisches Bundesamt, 2017). Hence, SLTs and other educational and health professions in Germany are frequently facing the challenge to assess Turkish-German bilingual children and to decide whether their speech (and language) development is progressing typically or atypically (Lüke & Ritterfeld, 2011). This task is particularly impeded for children's speech development since comprehensive phonology assessments to examine children's Turkish in this population, explicitly, are rare and lack internationally required linguistic and/or psychometric criteria (see Section 2.1.3). Additionally, information on Turkish-German bilingual children's typical phonological acquisition only exists from cross-sectional pilot studies (Salgert, Fricke, & Wells, 2012; Ünsal & Fox, 2002) which constrains generalisability and reliability. Also, atypical phonological development in Turkish-German bilinguals has only been explored in case studies (Fox-Boyer, Casper, & Önal, 2014; Tugay & Schultz-Ünsal, 2013). An internationally inconsistent use of criteria and markers for identifying speech sound disorders (SSD) in monolinguals, and especially in bilinguals, further complicates the situation (see Section 2.2.1). This demonstrates the need for comprehensive research in this field, which should aim to cover the following aspects:

(a) Design of a linguistically and psychometrically sound phonology assessment for Turkish-German bilingual children's Turkish abilities
(b) Collection of data on Turkish-German bilingual children's concurrent and longitudinal phonological abilities in both languages to identify typical development in this population
(c) Identification of potential clinical markers for detecting SSD in this population.

The present thesis presents a research project that aimed to address these three aspects. Following the design of a Turkish phonology assessment for Turkish-German bilingual children, a community sample of 84 Turkish-German bilingual children aged between 3;0 - 5;5 years was assessed on their phonological skills in both languages with single-word naming tests. About half of the children were followed up 12 - 15 months after the initial assessment and additionally tested with further quasi-language-independent psycholinguistic tasks. The gathered data lay the foundation for a description of Turkish-German bilingual children's phonological development as well as the identification of clinical markers to differentiate typical from atypical phonological acquisition in this population. For this, existing data on monolingual German- and Turkish-speaking children were additionally considered.

The structure of this thesis is as follows. Chapter 1 and 2 provide an overview of the sighted contemporary literature for this research. Chapter 1 addresses aspects of phonological acquisition (i.e., phonological theories, special characteristics of bilingual phonological acquisition, German and Turkish phonology, the phonological acquisition in typically developing monolingual German- and Turkish-speaking children), whereas Chapter 2 is dedicated to aspects regarding the differentiation of typical from atypical phonological acquisition (i.e., assessment of speech sound skills as well as the identification of bilingual children with suspected SSD).

Chapter 3 outlines the present research's design, provides information about participants and the recruitment process, and explains the assessment procedures including materials applied and scoring criteria set.

In Chapter 4, the results of the first data collection phase (t1) regarding children's phonological acquisition in Turkish and German are presented. This includes the acquisition of consonants, consonant clusters (CCs), speech accuracy as well as the production of phonological variations from adult-like speech. Further, potential influential factors are explored. These findings are subsequently discussed in Chapter 5.

Results of the follow-up assessment (t2) are presented in Chapter 6 and encompass the analysis of children's progress on consonant and CC acquisition as well as their progress on achieving speech accuracy and overcoming phonological variations. Outcomes of the follow-up assessment and the different developmental paths children demonstrated are discussed in Chapter 7.

Chapter 8 constitutes the last results chapter presenting the categorisation of children's phonological performances at both time-points as well as the analysis of potential psycholinguistic predictor variables for children's phonological performance in single-word naming tests. These findings are discussed in Chapter 9, which further addresses the clinical markers for SSD identified in the present research and compares the categorisation scheme applied with those frequently used in the literature.

The last chapter (Chapter 10) provides an overall discussion of the research aims of this thesis, an evaluation of the strengths and limitations of the present research as well as its clinical implications and future directions. The thesis closes with a summary and the conclusions.

1 LITERATURE REVIEW I – Acquisition of phonology

This chapter reviews the contemporary literature on the acquisition of phonology in monolingual and bilingual children and is structured into four sections. The first section provides an introduction into the variety of theories and models trying to explain the phonological acquisition and development in monolingual children with reference to bilinguals wherever possible. In Section 1.2, the specific characteristics of bilingual phonological acquisition as well as the factors influencing it are presented. The third section introduces the reader to the phonological systems of German and Turkish as well as their variations due to dialectal/accentual influences and cross-linguistic transfer in Turkish-German bilingual speakers. In the final section of this chapter, the current state of knowledge of the typical phonological acquisition in German and Turkish is presented which focuses on monolingual and bilingual children.

1.1 Theories in phonological acquisition

Questions such as *What abilities or mechanisms drive children's acquisition of phonology?; Which aspects of the phonological system (of their language) appear to be simple or challenging to acquire?; Does acquisition across languages follow similar developmental paths?* have attracted researchers' interests for about a century now. Whilst one of the initial approaches to understanding children's speech acquisition was formal linguistically motivated (cf. Jakobson, 1968), other disciplines (e.g., cognitive science/psychology, pedagogy, biology, second language acquisition) have increasingly contributed their knowledge to the understanding of the mechanisms of speech acquisition over the years and have highlighted the complexity of this process during children's development (Davis & Bedore, 2013; Vihman, 2014). As a result a multi-faceted perspective on speech acquisition has evolved which is strongly represented in the recent literature (Davis & Bedore, 2013).

In the following subsection, the most prominent streams of theoretical approaches to speech acquisition in monolinguals are outlined and compared. This is included for two reasons: (a) to provide a theoretical basis for the monolingual data to which the bilingual data of the present research are retrospectively compared and (b) since some phonological theories for monolinguals have been used to discuss and explain bilingual children's phonological data (e.g., Fabiano-Smith & Barlow, 2010; Fabiano-Smith & Goldstein, 2010a, 2010b; Procter, Bunta, & Aghara, 2015).

In order to acquire the phonology of their ambient language successfully, children need to perceive, process and store phonological knowledge and need to be capable of coordinating articulatory movements with their knowledge of the phonological system to produce intelligible linguistic output (Menn, Schmidt, & Nicholas, 2013). Further, children need to have the capacity to socially interact with their environment (i.e., through turn-taking, joint attention and intention-reading) to use the input they receive through their interlocutors for enhancing neural networks and mechanisms to encrypt and underpin phonological knowledge (Davis & Bedore, 2013; Menn et al., 2013). For this to be possible, linguistic input from the environment (e.g., through varying interlocutors) needs to be available from which the child can gradually learn how to employ their phonological knowledge about the ambient language appropriately to communicate their ideas about the world (Davis & Bedore, 2013). Considering these necessary abilities to master the acquisition of speech illustrates the complexity of language and its acquisition. A theoretical model aiming to explain the acquisition process thus needs to be able to account for:

a) the discrepancies of children's productions with the adult target (Barlow & Gierut, 1999)
b) the general patterns that can be observed across children and languages (Stoel-Gammon & Dunn, 1985)
c) the inter-individual variability as well as the variations that occur over time and which include progression, stagnation and regression in speech accuracy (Barlow & Gierut, 1999; Stackhouse & Wells, 1997; Stoel-Gammon & Dunn, 1985; Vihman, 2014)
d) the relationship between phonetic and phonological acquisition (Kent, 1984), the discrepancies that frequently occur between the two (Stoel-Gammon & Dunn, 1985) as well as how the child develops phonetic and phonological categories from the speech input (Vihman, 2014)
e) the role of input of the ambient language (Davis & Bedore, 2013; Stoel-Gammon & Dunn, 1985).

Moreover, a phonological acquisition theory needs to:

f) be compatible with cognitive and linguistic theories explaining learning as well as perception models to explain the relationship between perception and production (Ambridge & Lieven, 2013; Stoel-Gammon & Dunn, 1985)
g) incorporate biological and social-interactional aspects (e.g., attention) which provide children's physical and interactional prerequisites for speech acquisition (Kent, 1984; Vihman, 2014)
h) include testable and falsifiable predictions (Barlow & Gierut, 1999; Stoel-Gammon & Dunn, 1985).

Over the last few decades, several theoretical approaches have evolved, either in parallel or building on each other. Despite differing in detail their underlying core assumptions can be allocated to one of the following three approaches: formalist, perception and functionalist/emergentist approaches (Vihman, 2014). Within the scope of this research only theories accounting for phonological production are considered. Table 1.1 roughly chronologically presents the main theoretical approaches/models for phonological acquisition, which especially had an influence on the field of clinical phonology. Further, Table 1.1 displays their allocation to the two streams of theories that consider production, that is formalist (highlighted in grey) and functionalist (highlighted in blue), as well as their core hypotheses and main representatives. A juxtaposition of these two theoretical streams was preferred over a detailed presentation of each phonological model since aspects of both streams are recurrently discussed in research on bilingual phonological acquisition. This is the case as performances of bilingual children often seem to evidence the presence of both universal and language-specific structures and acquisition strategies (see e.g., Yavaş, 2015).

Table 1.1: Influential formalist and functionalist phonological (acquisition) theories

Theory	Core statements/hypotheses	Main representative(s)[1]
Structuralist model	• Universal principles guide development • Acquisition of phonological oppositions (unmarked contrasts are acquired before marked contrasts) has a universal character • Babbling is random and unrelated to speech	Jakobson (1968)
Generative models Including: Generative Phonology Natural Phonology Nonlinear Phonology Prosodic Morphology Sonority Sequencing Principle	• Children start their speech production with a set of universal mental representations (principles/rules/processes) and a limited number of parameters that allow for cross-linguistic variations (to be learned) • Children's underlying phonological representations are adult-like • Language experience prompts innate knowledge and determines language-specific parameters → phonological rules/processes need to be unlearned, constrained and ordered • Phonological acquisition is not a process of conscious learning	Chomsky & Halle (1968), Donegan & Stampe (1979), Clements (1990), Goldsmith (1990), Bernhardt (1992), McCarthy & Prince (1998)
Optimality Theory[2]	• Language is a system of conflicting universal constraints (i.e., markedness and faithfulness constraints) • Speech acquisition proceeds by re-ranking the constraints in a language-specific hierarchy • Observed 'surface' forms in child output and in languages arise from the resolution of conflicts between these ranked grammatical constraints	Barlow & Gierut (1999), Stemberger & Bernhardt (1999) Prince & Smolensky (2004)

[2] Some Optimality Theory approaches are strongly based on Generative Phonology principles whereas others are not and almost hold an emergentist view (Stemberger & Bernhardt, 1999). Therefore, this theory is presented separately.

Theory	Core statements/hypotheses	Main representative(s)[1]
Behaviourist models	• Ability to learn is innate and triggers phonological acquisition • Babbling and phonological knowledge of the ambient language is shaped through imitation and external reinforcement (i.e., behaviour modification) to conform to adult-like speech	Mowrer (1952), Skinner (1957), Olmsted (1966)
Biologically based models Including: Self-organising models Articulatory Phonology Dynamic Systems Theory Frame/Content Theory Lindblom's model	• No a priori phonological knowledge is required • The use of the vocal system is driven by the interaction between the parent and the child • Phonological patterns in late babbling resemble those of early meaningful speech as both are universal and controlled by phonetic tendencies • Individual differences are permitted by an interaction of phonetic and cognitive factors and create opportunities for change and adaptation	Locke (1983), Kent (1984), Browman & Goldstein (1992), Lindblom (1992), Thelen & Smith (1994), MacNeilage (1998)
Cognitive/usage-based models Including: Neighbourhood Density Whole-word Phonology Linked-Attractor model Connectionist neural net model	• Child formulates and tests hypotheses about the phonological system being acquired • Linguistic (phonological) system is constantly altered while being utilised → phonological representation is emergent • Type and token frequency in the input shape phonological structure • Commonly occurring patterns in children's speech are ascribed to the universal physiology of children's articulatory, auditory and central nervous systems • Language learning is associated with other forms of learning	Ferguson & Farwell (1975), Macken & Ferguson (1983), Vihman (1996), Bybee (2001), Pierrehumbert (2003), Elman (2005), Storkel (2006), Menn, Schmidt & Nicholas (2013)
Emergentist model	• Linguistic knowledge emerges through the interaction and interconnectivity of child-internal general-purpose capacities and mechanisms (i.e., production, perception, neural-cognition, social interaction) with social and cultural input from the environment	Davis & Bedore (2013)

Note: Formalist approaches, functionalist/emergentist approaches, [1]This list of authors is not exhaustive.

The shared basic assumption of formalist approaches is that language (incl. phonological) acquisition is guided by a universal (i.e., valid for all languages) set of principles and parameters, rules or constraints (i.e., representations of universal grammar). Children are considered to be born with this abstract linguistic knowledge which unfolds with language experience and leads to a largely invariable acquisition process across languages (Chomsky & Halle, 1968; Jakobson, 1968). Jakobson (1968) asserted that the regularities that guide the structure of adult languages are not only very similar across languages but also set the order of phoneme acquisition in children, from simple to complex and by focusing on maximally different feature contrasts of speech sounds. As a consequence, linguistic features (e.g., sounds) that are shared across most languages are assumed to be unmarked (i.e., the default version) and acquired before those that are (relatively) language-specific and marked (Reimers, 2015; Yavaş &

Topbaş, 2004). This emphasises the pure linguistic perspective on speech acquisition, which is generally held by formalist approaches, and the minimisation of influences from other capabilities such as, for example, cognition and social interaction (Ambridge & Lieven, 2013).

Despite the shared[3] view on innate mental representations of language knowledge, the different formalist approaches have in part very distinct views on the role of the ambient language input in the acquisition process. In Optimality Theory approaches (and Nonlinear Phonology), for example, the child is considered to be stimulated by the frequency of certain phonological features in the input to re-rank universal (markedness) constraints to adjust the output to the specificities of the ambient language (Davis & Bedore, 2013; Stemberger & Bernhardt, 1999). Markedness constraints are thus substituted by faithfulness constraints and therefore allow for cross-linguistic variation (Prince & Smolensky, 2004). In contrast, Natural Phonology approaches, for example, hypothesise that children have accurately perceived and stored the phonological information of their ambient phonological system (by phonological rules) before they start their speech production (Donegan & Stampe, 1979).

Additionally, formalist approaches generally disregard the role of the cultural context in which phonological input occurs, the influence of the social and interactional skills of the child necessary to communicate effectively, as well as the child's physical capacities required for a successful acquisition of phonological systems (Davis & Bedore, 2013). Therefore, they are hardly in line with other learning theories. Furthermore, some formalist approaches (esp.: Structuralism, Generative Phonology, Natural Phonology) have difficulties in explaining individual differences across children as well as the U-shaped learning process of speech acquisition. The latter is known from psychological research (e.g., Werker, Hall, & Fais, 2004) and describes the transient reorganisation of children's phonological system or the emergence of an organised system (Vihman, 2014) triggered by the acquisition of new competences. These new competences are considered to change the way children used to interpret the input of their ambient language(s) and result in an apparent temporary loss of skills which are regained at a later point in time (Werker et al., 2004). Formalists, however, assume the acquisition process to be mainly linear, guided by universals and similar across children which is often not in line with children's actual performances (Stoel-Gammon & Dunn, 1985).

In contrast, (most) functionalist approaches share the emphasis on the active role children take in acquiring speech by interacting with their environment and generating

[3] It has to be acknowledged that some Optimality Theory approaches do not adhere to the innateness of children's mental representations (constraints) but see them as based in acoustic/phonetic features (Boersma, 1998; Stemberger & Bernhardt, 1999).

their own hypotheses about the ambient phonology from the input they receive. This then helps them to tailor their phonological output accordingly (Bybee, 2001; Vihman, 2014). During the expansion of their lexicon children increasingly built up their own phonemic systems (Bybee, 2001) and the pace of this development is considered to be constrained by the phonetic (Kent, 1984) and processing abilities (Pierrehumbert, 2003) at any given time. This is considered to explain the presence of individual differences in the strategies used to acquire phonological knowledge and the various phonological output forms produced within and between children as well as across languages (Davis & Bedore, 2013). Children are not considered to have incomplete/immature adult-like phonological systems but a phonological system in its own right (Vihman, 2014).

Further, the entire speech acquisition process is considered to be self-organisational (MacNeilage, 1998) and of an emerging nature. Functionalist approaches generally observe and interpret the phonological development in children from an interdisciplinary perspective, incorporating insights from psycholinguistics (i.e., cognitive), biology, perception, social interaction, neuro-cognition etc. Despite these agreements, the individual approaches categorised as functionalists can be differentiated (but not only) by the emphasis they give to the outlined disciplines. Behaviourist models highlight the child's interaction with the environment (reinforcement; e.g., Olmsted, 1966; Skinner, 1957); biologically-based models stress the importance of the physical articulatory/motor speech abilities for speech acquisition but also acknowledge the role of cognitive and social-interactional components (e.g., Kent, 1984; Locke, 1983); and cognitive/usage-based theories focus on the interaction of children's cognitive capabilities (i.e., processing and knowledge, hypothesis building) but additionally recognise the role of perception, social interaction, motor abilities and ambient language input (i.e., frequency) (e.g., Bybee, 2001; Menn et al., 2013). The emergentist model, however, assumes that child-internal biological and social-interactional capacities engage with linguistic and cultural input from interlocutors as well as with each other in a two-way fashion so that one isolated aspect does not have a causal effect on children's phonological acquisition (Davis & Bedore, 2013). Hence, the emergentist model incorporates all disciplines addressed by the individual functionalist (and also some formalist) approaches, expands on them and imputes the child-internal abilities and mechanisms a strong interplay during development from which phonological knowledge and behaviour can emerge. With this it is intended to embed speech acquisition research into the study of other natural phenomena (Davis & Bedore, 2013). The strong acknowledgement of the ambient language phonology, interactional instinct and cultural input provides the opportunity to also explain bilingual children's acquisition within this approach (Iglesias & Rojas, 2012).

In conclusion, although formalist and functionalist theoretical approaches to children's speech acquisition have similar intentions (e.g., describe the mismatch between children's and adult's productions, explain general speech patterns observed across languages) their understanding of which skills and mechanisms are involved in children's acquisition and which internal and external factors may have an influential role on the speech outcome is crucially different, especially regarding:

- the driving forces of speech acquisition (i.e., innate vs. learned/emerging)
- the acquisition of linguistic perception
- the role of ambient language input
- the degree to which they can explain individual variability
- the integration of further aspects and functions of language as well as child-internal physical and social skills (Davis & Bedore, 2013; Stoel-Gammon & Dunn, 1985).

Table 1.2 shows the degree to which each formalist and functionalist approach addresses/covers the criteria necessary for a phonological acquisition theory (as outlined at the start of this section).

Table 1.2: Overview of the phonological acquisition aspects the discussed theories can account for/consider

	Aspects to be considered/accounted for by a theory							
Theory	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**
Structuralist model	+/-	+	-	-	-	-	-	-
Generative models	+	+/-	+/-	-	+/-	-	-	+/-
Optimality Theory	+	+	+/-	-	+/-	-	-	+
Behaviourist models	+/-	+/-	-	+/-	+/-	-	+	+/-
Biological models	+	+	+/-	+/-	+/-	+	+	+
Cognitive/Usage-based models	+	+/-	+	+/-	+	+	+/-	+/-
Emergent model	+	+	+	+	+	+	+	+

Note: **a**: mismatches between children's and adult-like speech, **b**: general speech patterns observable across children and languages, **c**: inter- and intra-variability during development, **d**: relationship between phonetics and phonology, **e**: role of ambient-language input, **f**: compatibility with other learning and perception theories, **g**: inclusion of biological and social-interactional aspects, **h**: testable and falsifiable hypotheses, formalist approaches, functionalist approaches, **+**: is accounted for, **+/-**: is accounted for in some approaches of this theory/to some extent within this theory, **-**: is not accounted for/not addressed

To some extent, formalist approaches appear to be able to explain the linguistic surface patterns that describe the mismatches between adult-like and children's speech as well as the general patterns evident across languages. However, they demonstrate weaknesses in the ability to sufficiently explain individual variation (apart from Nonlinear Phonology and Optimality Theory), in considering the role of ambient language input (although some Generative Phonology and Optimality Approaches try to account for this), and in being compatible with other theories (Stoel-Gammon & Dunn, 1985). Functionalist theories, in contrast, provide predominantly satisfactory explanations for the patterns found in children's speech output, children's variations from the adult target as well as its variability across languages and children (Vogel Sosa & Bybee, 2011). The majority considers the relationship between phonetic and phonological aspects and addresses children's acquisition of the differences between them. Further the consideration of different learning abilities as well as child-internal capabilities is strongly incorporated in most functionalist approaches (Ferguson & Farwell, 1975; Menn et al., 2013; Vihman, 1996). Thus, functionalist theories meet the majority if not all (e.g., emergentist model) of the aspects they expect a sound phonological theory to explain in children's speech acquisition.

While this is not a theoretical thesis which aims to evaluate or confirm/disprove a specific theory, this theoretical background will be used as a reference point for the remainder of the literature review and picked up again in the discussion of the data of this research. The specific characteristics of bilingual phonological development as well as some frequently cited theoretical models aiming to explain it are presented and discussed in the next section.

1.2 Phonological acquisition in bilingual children

The difference between monolingual and bilingual speech acquisition is the bilingual child's need to acquire two phonological systems, including the language-specific phonetic-phonological features (e.g., phonotactic restrictions and stress patterns) in parallel (simultaneous acquisition)[4] or shortly after one another (sequential acquisition). These two systems are composed of shared and unshared phonological features whose presence in the bilingual's input is highly variable due to different language input and output patterns among children. In order to understand and employ the languages accurately it is the bilingual child's challenge to map the language-specific phonological features to the correct languages (Davis & Bedore, 2013). One of the major aspects that have attracted researchers' interests is the question of whether bilingual children use one shared or two separate phonological systems to store and process their knowledge

[4] The terms used to classify children who acquire two (or more) languages regarding their AoA of their languages vary considerably in the literature (cf. De Houwer, 2009; Kohnert, 2013; Meisel, 2004; Paradis, Genesee, & Crago, 2011). Traditionally, the terms *simultaneous* and *sequential acquisition* are used to refer to children starting to acquire both languages from birth (or during their first three years of life) and those starting to acquire one language from birth and the other at a later point in time (usually after the age of three), respectively (Kohnert, 2013; Yip, 2013). These terms are also used throughout this thesis – also in cases where the original source included a different term but which referred to a similar AoA.

of the two languages. And if they used separate systems at what age this differentiation becomes evident.

In addition, acquiring two languages has been proposed to require more processing and memory capacities as the language input children receive is more variable (i.e., input from two linguistic systems; Iglesias & Rojas, 2012) which may quantitatively and qualitatively affect the acquisition process and language competences (Michael & Gollan, 2005; Paradis & Genesee, 1996). Thus, it has been investigated whether being exposed to two languages during early childhood has a delaying or an accelerating effect on the acquisition of phonological systems (e.g., Fabiano-Smith & Goldstein, 2010b; Paradis, 2001; Paradis & Genesee, 1996). Further, bilinguals provide an ideal population, in comparison to monolinguals, to explore which socio-linguistic variables may have an influence on the rate and quality of phonological acquisition.

The findings reported regarding these questions are outlined and discussed in this section using the following structure:

1) Shared vs. separated phonological systems
2) Rate of acquisition
3) Quality of acquisition
4) Factors influencing phonological acquisition.

For all subsequent discussions it should be considered that the specific language combination of a bilingual is likely to affect qualitative aspects of speech acquisition since languages vary considerably in their linguistic (i.e., also phonological) structures (Yip, 2013). Thus, all comparisons and generalisations should be interpreted with caution and reference to the languages involved.

1.2.1 Do bilingual children use shared or separated phonological systems?

It is well-known from perception studies that bilingual children are able to distinguish the phonologies of their ambient languages from an age as early as the babbling stage (Bosch & Sebastian-Galles, 2003; Byers-Heinlein, Burns, & Werker, 2010; Pons, Bosch, & Lewkowicz, 2015). Paradoxically, studies examining children’s early lexical utterances have revealed little consensus on whether bilingual children acquire and use one shared or two separated phonological systems. Investigating phonetic inventories, Schnitzer and Krasinski (1994, 1996) reported different acquisition paths of two Spanish-English bilingual children aged 1;1 and 1;6 years at t1 respectively. In the first case, the child used identical substitutions of single phones in both languages and initially exhibited similar acquisition rates for shared sounds. The authors considered this to be evidence for the phones being represented in a single phonological system which separated into two systems as soon as the child had acquired the ability to consistently produce the fine phonetic differences required for the target sounds in both Spanish and English (Schnitzer & Krasinski, 1994). In the second case, however, the child demonstrated an overall slower but independent acquisition of the individual phones of his languages. Phonetic targets were consistently produced very accurately and in the target language which the authors interpreted as a support for the presence of two separated phonological systems (Schnitzer & Krasinski, 1996). Similarly, other researchers interpreted the dissimilar phonological performances in a bilingual's two languages as evidence for two separated phonological systems and predominantly rejected the idea of a single phonological system. The distinctions children exhibited across their languages included babbling patterns that were adjusted to the language of the interlocutor (Maneva & Genesee, 2002; Yang & Zhu, 2010), the acquisition of shared sounds in one language before the other (Montanari, 2011), the application of the correct stress pattern in each language (Keshavarz & Ingram, 2002), the variation of phonological realisations of certain phonemes across languages (e.g., a child may have difficulties with the realisation of /v/ in both languages but stops it in language A (/v/ → [b]) and affricates and devoices it in language B (/v/ → [p͜f]) (Anderson, 2004), the adherence to language-specific phonotactic rules (Holm & Dodd, 1999c; Paradis, 2001) and the use of phonological patterns[5] in only one language although they could have occurred in both languages (Holm, 2007).

Despite these findings, researchers have argued that simply because children distinguish between the phonologies of their two languages this does not necessarily imply that their phonologies are represented in two separate systems (Johnson & Lancaster, 1998; Vihman, 2002). Further, Paradis (2001) argues that it is difficult to identify whether similarities between a bilingual's phonologies occur because of a shared phonological system or due to the absence of language-specific features at that developmental stage, which would also be present in monolingual children. This demonstrates the difficulty with testing the hypothesis for distinct phonological systems. In the view of an emergentist approach, Vihman (2016) proposes that the question of one system or two does not need to be posed at all since a child would individually construct the phonological characteristics, structures and rules of their ambient language based on their respective input and item learning. The Estonian-English and German-

[5] Phonological patterns (also labelled as error patterns or phonological processes in previous work) are a measure used to describe the phonological system of a child in relation to adult realisations of a target word. They reflect a general tendency of production that consistently deviates from the adult target and usually affects a group of phonemes (Dodd, Holm, Zhu, & Crosbie, 2003). In this research the term *phonological patterns* will be used throughout to avoid theoretical assumptions associated with the term *phonological processes* and to avoid using the term *error* in developmental speech.

English participants in her study, for instance, demonstrated individual rather than universal word templates and showed no stable differentiation between the word templates used for productions in their two languages. Instead their productions frequently exhibited influence by their ambient languages (Vihman, 2016).

This overview highlights how the diverse theoretical assumptions guiding the interpretations of findings complicate reaching a consensus with regard to the existence of shared or separated phonological systems in bilingual children.

1.2.2 Rate of phonological acquisition

Although both monolingual and bilingual phonological acquisition are amongst others influenced by the maturation of children’s motor abilities and milestones (from first cries to full sentences) are reached at similar ages (De Houwer, 2009), the rates at which bilingual and monolingual children acquire fine phonological features (e.g., phonetic contrasts) expressively may differ (Davis & Bedore, 2013). Results of previous studies with bilinguals revealed a faster or equally rapid acquisition of phonological output skills (e.g., Gildersleeve-Neumann & Wright, 2010; Goldstein, Fabiano, & Washington, 2005; Goldstein & Washington, 2001; Grech & Dodd, 2008; Lin & Johnson, 2010; Lleó, Kuchenbrandt, Kehoe, & Trujillo, 2003; Montanari, 2011; So & Leung, 2006) as well as incidences of delayed acquisition rates when compared to typically developing monolinguals (Bunta, Fabiano-Smith, Goldstein, & Ingram, 2009; Fabiano-Smith & Goldstein, 2010b; Gildersleeve-Neumann, Kester, Davis, & Peña, 2008; Goldstein & Washington, 2001; Holm & Dodd, 2006). Thus, bilingual children’s rate of speech acquisition was not found to be homogeneously decelerated or accelerated. These differences and similarities to the rate of monolingual acquisition apply to a variety of phonetic and phonological skills such as consonant and vowel accuracy, acquisition of specific syllable structures (e.g., coda-productions), percentage of occurred phonological patterns and the age of suppression of phonological patterns. To account for the different outcomes in previous studies and to define how bilingual children acquire the phonological systems of their languages, researchers have formulated a series of hypotheses. Paradis and Genesee (1996) as well as Paradis (2001), for example, who did not specifically target phonology, postulated that bilingual children acquire two separate linguistic systems which are likely to systemically interact with each other (*Interdependence Hypothesis*). This interaction, which is theoretically embedded in formalist assumptions (i.e., generative phonology), may exhibit itself in three different ways (applied to phonology here):

1) *Transfer* – using unique phonetic-phonological features of one language in the other (see also Section 1.2.3.3).
2) *Acceleration* – phonological properties that are acquired early in one language may have an accelerating effect on the acquisition of these properties in the other language and are consequently earlier acquired in bilinguals than in monolinguals.
3) *Deceleration* – the processing and storing load while acquiring two languages in parallel may decelerate the acquisition resulting in a slower phonological acquisition in bilinguals than in monolinguals.

Although, the studies by Paradis and Genesee (1996) and Paradis (2001) did not reveal any clear support for the interdependence hypothesis, Fabiano-Smith and Goldstein's (2010b) investigation yielded evidence for the aspects of transfer and deceleration of the interdependence hypothesis in Spanish-English bilinguals. The performances of their participants did not evidence the aspect of acceleration in its strict sense but they fell within the normal range of monolinguals for some phonological measures. Fabiano-Smith and Goldstein (2010b) considered a similar rate of acquisition in bilingual and monolingual children as a variation of acceleration since achieved knowledge of phonological properties in one language may support the acquisition of these or related properties in the other language. The presence of transfer, (variation of an) acceleration and deceleration partly depended on the phonetic-phonological aspect under study and varied across the two languages. For example, Fabiano-Smith and Goldstein's (2010b) bilingual participants demonstrated a decelerated phonological acquisition of fricatives in both languages but had acquired Spanish stops and nasals as well as English nasals, affricates and glides within the normal range of monolinguals, hence evidencing a variation of acceleration as well. Despite not specifically geared to the interdependence hypothesis, results of other studies have confirmed a simultaneous presence of accelerated and decelerated aspects in bilingual children's development if compared to monolingual peers (Goldstein & Washington, 2001; Morrow, Goldstein, Gilhool, & Paradis, 2014; So & Leung, 2006).

Other researchers consider mechanisms of first language (L1) learning as a division of the processes required for second language (L2) learning and bilingualism (MacWhinney, 2005a). With the *Unified Competition Model*, MacWhinney (2005a, 2005b, 2008) describes a model that theoretically encompasses the mechanisms behind L1 acquisition, L2 acquisition and bilingualism. The model's premise is that the language input bilingual children receive is distributed across both of their languages and thus does not resemble the amount of input per language found for monolinguals. Consequently, features that are shared across languages occur to a proportionally higher frequency in a bilingual's input than in a monolingual's input. According to MacWhinney

(2005b), bilingual children are specifically sensitive to the shared aspects, which form strongly reliable cues, which enable the children to use their knowledge and abilities in language A for more precise realisations in language B. Due to this, their phonological skills in the shared properties are likely to turn out to match or be even more advanced than those of monolinguals (referred to as *positive transfer* in the model; MacWhinney, 2008). Conversely, language-specific elements (e.g., unshared phones) are assumed to be relatively infrequent in the bilingual's input as they only occur in one language for which the general amount of input is lower in bilinguals than in monolinguals. Therefore, language-specific features are proposed to be less frequently and strongly perceived by bilinguals (less reliable cues) so that they develop less skills in producing them which is likely to result in a decelerating effect on the rate of speech acquisition (referred to as *negative transfer* in the model; MacWhinney, 2008). *Negative transfer* also includes the incorporation of grammatical features in one language from the other (MacWhinney, 2005a) which is labelled as *transfer* in Paradis & Genesee's (1996) hypothesis. Hence, the model is in line with functionalist theories (esp. connectionist) emphasising the significant role of ambient language input and self-organisation for the process of speech acquisition (see Section 1.1).

For the *Unified Systems Model* to be unbiased, it is important to control for children's language use and proficiency as these two factors may vary across bilinguals and have an influence on their phonological performance (Goldstein & Bunta, 2012, see further discussion on this in Section 1.2.4). In their study, Goldstein and Bunta (2012) found evidence for both of MacWhinney's proposed transfer effects while having language use and proficiency controlled. Their five- to six-year-old monolingual Spanish-speaking children significantly outperformed Spanish-English bilingual peers on the measure of percentage of consonants correct (PCC) for stops, indicating a *negative transfer*. Regarding the percentage of produced phonological error patterns, though, bilingual children exhibited equal or even lower percentages than their monolingual English-speaking peers, thus demonstrating a *positive transfer*. Even if language use and proficiency are controlled for, the model might be biased by other aspects. The fact that shared properties between two languages exist, for example, does not necessarily imply that they occur to equal frequency in both languages. Considering the language combination of Turkish and German, it can be noticed that despite both phonological systems share the phoneme /n/, its frequency of occurrence in German is nearly twice as high as in Turkish (Trost, 2007-2014). Also, the degree of phonological saliency for a shared phone may differ for the respective languages (see Section 1.4.1 for an example of the potential effect of phonological saliency on the acquisition of German) and may especially be different in the input bilingual children receive. Thus, the rate of speech acquisition in bilinguals seems to differ from the average rate of acquisition in monolinguals and appears to be attributable to a variety of aspects.

1.2.3 Quality of phonological acquisition

Despite the described quantitative differences between monolingual and bilingual children, research has shown that bilingual children broadly follow a similar acquisition path as monolinguals do. That is, children acquire simple before complex structures (e.g., consonants before CCs; Fabiano-Smith & Barlow, 2010) and show improvements in their phonological skills with age (e.g., Grech & Dodd, 2008; Lee, Ballard, & Purdy, 2015b; Morrow et al., 2014). Additionally, simultaneous bilinguals were found to predominantly abide by the phonotactic restrictions of each of their languages from an early age (Keshavarz & Ingram, 2002). This similar acquisition path, however, may not be followed for both of their languages (Fabiano-Smith & Barlow, 2010) due to, amongst others, an influence of children's language experience patterns (see Section 1.2.4). These and further qualitative differences to monolingual acquisition are presented in the following subsections.

1.2.3.1 Distributed phonological skills across languages

The level of language proficiency bilingual children may achieve in each of their languages usually differs across their languages, since a bilingual is not a monolingual times two (Grosjean, 1989). These differences may become apparent quantitatively (e.g., by achieving significantly lower PCC-scores in language A than in language B), or present themselves qualitatively (e.g., by demonstrating phonological patterns in only one language although they could occur in both languages). Investigating the speech of eight three-year-old bilingual Spanish-English-speaking children and their monolingual peers with whole-word measures, Bunta et al. (2009) discovered lower phonological mean length of utterance (pMLU) scores in the bilinguals' English than in their Spanish speech output. The PCC-scores, however, were higher for the bilinguals' English productions than for their Spanish. Similar findings were reported in Holm and Dodd (2006) who examined the consonant accuracy of 40 Cantonese-English bilingual children aged 2;2 - 5;7 years in both languages. Participants exhibited a substantially lower consonant accuracy in English than in their L1 (i.e., Cantonese). Furthermore, quantitative differences across languages were also reported regarding the overall number of different phonological patterns types (Holm & Dodd, 1999c, 2006). Studies that reported qualitative differences across languages mentioned dissimilar types of produced phonological error patterns in the bilingual's languages (e.g., Anderson, 2004;

Kim, Ballard, & McCann, 2016). These appeared, for example, in the form of contradictory patterns such as backing of alveolar stops in Cantonese but fronting of velar stops in English (Holm & Dodd, 2006), in the form of the same pattern across languages but which affected different sounds such as stopping of fricatives in one language but stopping of nasals in the other (Holm, Dodd, Stow, & Pert, 1999), or in the form that sounds are affected by the same pattern in both languages but to a different frequency (Brice, Carson, & O'Brien, 2009). Altogether, these findings support Grosjean’s (1989) statement that one cannot necessarily assume equal competence in a bilingual’s two languages, but rather a qualitatively and quantitatively distributed phonological proficiency.

1.2.3.2 Differences in the observed phonological pattern types compared to monolinguals

A further aspect of difference between monolingual and bilingual children is the type of phonological patterns that occur in bilingual children’s speech. Across studies, the majority of phonological patterns present in bilinguals was found to resemble those of younger or same-aged monolingual children of each language (e.g., Hack, Marinova-Todd, & Bernhardt, 2012; Holm & Dodd, 2006; Preston & Seki, 2011; Salameh, Nettelbladt, & Norlin, 2003; So & Leung, 2006). However, some studies reported a subset of the phonological patterns to be considered as unusual for monolingual peers (Dodd, So, & Li, 1996; Holm & Dodd, 1999c, 2006). For example, the Cantonese-English bilingual children in Holm and Dodd’s (2006) study exhibited phonological pattern types that were regarded as uncommon for monolingual peers; one half of them for both languages and the other half for only one of their languages. In a recent study by Kim et al. (2016), the proportion of phonological patterns considered to be unusual for monolingual Korean- or English-speaking children was low compared to those being age-appropriate or delayed but still encompassed 13.5% and 17.3% respectively. This is a non-negligible amount of phonological patterns that might lead SLT’s to misinterpret those as clinically atypical patterns.

In addition, (retrospective) comparisons to monolingual children have revealed that bilingual children are regularly less accurate in their realisations of vowels in each language (e.g., Gildersleeve-Neumann, Peña, Davis, & Kester, 2009; Gildersleeve-Neumann & Wright, 2010; Holm & Dodd, 2006; Johnson & Lancaster, 1998). This is interesting since vowels are usually acquired very early in typically developing monolinguals and vowel errors are rather scarce, even in children with phonological disorders (Donegan, 2013; Goldstein & Pollock, 2004). Gildersleeve-Neumann et al.

(2009) explain this phenomenon by the different complexities of vowel systems across languages which might cause confusions during phonological development.

Further analyses of especially those pattern types that were unusual for monolingual development, however, frequently revealed an influence from the respective other language of the bilingual child (e.g., Kim et al., 2016; Preston & Seki, 2011). The characteristics of this interaction between two languages is outlined and discussed in the following section.

1.2.3.3 Cross-linguistic transfer

Languages in contact, whether within an individual or within a group of different speakers, are prone to be influenced by each other (Backus, 2006; Backus, Jørgensen, & Pfaff, 2010; Paradis & Genesee, 1996). Interactions between languages within a speaker are frequently referred to as *(cross-linguistic) transfer* (or *negative transfer* in MacWhinney's (2005a) model). These special characteristics in bilingual children's speech may occur on both the segmental and the suprasegmental level. A segmental cross-linguistic transfer (CLT) describes the use of a language-specific sound in the productions of the other language (e.g., the German sound /ʁ/ is substituted by the Turkish equivalent /r/: *rot* (red) /ʁo:t/ → [ro:t]) (Fabiano-Smith & Goldstein, 2010b; Paradis & Genesee, 1996). Support for the presence of segmental CLT comes from a large body of studies including a variation of different language combinations (e.g., Anderson, 2004; Fabiano-Smith & Barlow, 2010; Fabiano-Smith & Goldstein, 2010b; Gildersleeve-Neumann et al., 2008; Gildersleeve-Neumann et al., 2009; Holm & Dodd, 1999c; Salameh et al., 2003). CLT on the suprasegmental level refer to the use of phonological patterns or phonotactic restrictions of one language in the productions of the other. Kim et al. (2016), for example, reported the insertion of a vowel into English initial consonant clusters for some of their 52 Korean-English bilinguals. This pattern is explainable by the syllable structure of Korean which does not include consonant clusters. Vowel epentheses in consonant clusters form a very frequent type of suprasegmental CLT having been reported for Pakistani-Heritage speakers of English aged 4;8 to 7;5 years (Holm et al., 1999), Mandarin-English bilingual children aged between 5;0 and 7;11 years (Lee et al., 2015b), a Spanish-Mandarin-Taiwanese trilingual child assessed from the age of 1;3 to 2;0 years (Yang & Zhu, 2010), and even for adult second language speakers (Chilla, Rothweiler, & Babur, 2010).

Other transferred phonological patterns from phonotactic restrictions in one of a bilingual's languages include a later acquisition of the Spanish rule for voiced stop spirantisation in three Spanish-German bilinguals aged between 1;3 and 4;0 years (Lleó & Rakow, 2005), palatalization of English consonants in 14 3;3 - 5;7-year-old Russian-

English bilinguals (Gildersleeve-Neumann & Wright, 2010), the substitution of the German /ʁ/ by [l] in the speech acquisition of five typically developing Russian-German bilinguals aged 3;0 to 4;8 years (Büttner, 2012), the use of [l] to replace /ɹ/ in the English productions of a four-year-old Russian-English bilingual child in Anderson (2004), as well as word-final stop aspiration, lateralisation of flap and laxing in the Korean-productions of 3;0 - 7;11year-old typically developing Korean-English bilinguals (Kim et al., 2016). The list of possible types of CLT in bilingual speakers is unlimited and dependent on the respective phonological systems in contact. The identification of CLT is not always clear cut as developmental and cross-linguistic transfer effects may operate at the same time reflecting underspecified realisation rules because of the existence of two phonological systems that are being restructured (Gildersleeve-Neumann et al., 2008). This became especially apparent by the inconsistent occurrence of fronting of /ʃ/ in Korean-English bilingual's English productions (Kim et al., 2016). In Korean phonology, the appearance of /s/ and /ʃ/ is distributive whereas in English it is contrastive.

As demonstrated by these examples, CLT may occur in one or both directions, that is from L1 to L2 and/or the reverse. According to Keshavarz and Ingram (2002) as well as So and Leung (2006), the transfer direction is dependent on the language the child is exposed to the most. That is, the more a child hears and uses language A the more likely (s)he will be transferring language-specific features to language B. It is, however, also conceivable that the context in which a child uses a certain language may influence the occurrence of CLT (e.g., if a child is asked to speak their L1 in a typical L2 setting; Fabiano & Goldstein, 2005). Overall, occurrence of CLT in bilingual children was found to be very low (Fabiano-Smith & Goldstein, 2010b; Fabiano & Goldstein, 2005; Gildersleeve-Neumann et al., 2008; Gildersleeve-Neumann et al., 2009; Goldstein et al., 2005; Keshavarz & Ingram, 2002; Schnitzer & Krasinski, 1994, 1996) and mainly of dynamic nature. That is, in contrast to a foreign accent (static transfer), their occurrence seems to be rather related to a break down in bilinguals encoding mechanisms because of tiredness or stress (see Grosjean, 2013). In clinical settings CLT may increase the risk of misinterpreting a child's speech abilities, especially in cases where the examining SLT is not informed about the differences in the phonological systems of the child's languages (Fabiano & Goldstein, 2005; Preston & Seki, 2011).

1.2.3.4 Variability of phonological skills over time

Phonological acquisition in monolingual children is, similar to the development of other language domains, guided by maturation. Thus, with increasing age, children's phonological skills improve until they are acquired to adult standards (McLeod, 2013; Saaristo-Helin, 2009; Weiss, 2007). In general, the same age-related effect was found for bilingual children (Gildersleeve-Neumann et al., 2008; MacLeod, Laukys, & Rvachew, 2011; Morrow et al., 2014). Independent of the respective language combination bilingual children improved their phonological skills on various aspects (e.g., speech accuracy, percentage of occurrence of phonological patterns). However, some longitudinal investigations of bilingual children's phonological development revealed a significant variability in children's performances. Three of Anderson's (2004) five bilingual participants aged 3;9 - 4;9 at t1 who were assessed five times in an interval of one to two months, for example, did not produce higher PCC-scores over time. Instead they demonstrated stable PCC-scores in both languages with the English ones being below 90%. Also, Gildersleeve-Neumann et al.'s (2009) six Spanish-English bilinguals aged 3;2 - 3;10 years at initial assessment demonstrated a considerable inter-child variability in the frequency of produced phonological patterns. This included children demonstrating decreases, stabilisations and increases over time (i.e., within 8 months). Whereas variability and a U-shaped developmental course in the phonological performances of (bilingual) children aged younger than three years (as observed in the longitudinal investigations by Keshavarz and Ingram, 2002, and Holm and Dodd, 1999c) may be expected due to children's emerging linguistic systems (Vihman, 2014), a reorganisation of phonological systems was not considered to be the case for bilinguals of the age ranges examined by Gildersleeve-Neumann et al. (2009) and Anderson (2004) and the intervals in which children's performances were assessed in these studies. However, based on the longitudinal performances of 16 3;0- to 5;11-year-old Korean-English bilinguals Kim, Ballard and McCann (2017) propose that a U-shaped developmental course may occur more frequently or persist longer in bilingual children due to their necessity to specify phonemes and phonological realisation rules not only within one language but also between their languages. This implies that children's information processing systems may be overloaded so that they fall back on a level of information processing they used to apply before while restructuring their phonological systems (Werker et al., 2004). Further, in contrast to the relatively stable context of phonological acquisition in monolinguals, the circumstances under which bilingual children acquire their two languages may change with time (e.g., due to enrolment in school). With this, language exposure and usage patterns as well as the importance for maintaining proficiency in both languages may alter from one point in time to another. Thus, it has been suggested that variability in phonological performance may have resulted from a change in children's language experience patterns (Gildersleeve-Neumann et al., 2009). To what extent socio-linguistic variables might influence phonological acquisition in bilinguals is outlined in the following sections.

1.2.4 Factors influencing the phonological acquisition in bilingual children

In context of the heterogeneity of the bilingual population it becomes apparent that the number of factors potentially impacting on a child's phonological acquisition is virtually unlimited. The acquisition context of individual children including the age of acquisition of the L1 and L2, the social setting of language learning and language use, the status and acceptance of each language in society as well as language dominance patterns vary considerably across children (MacLeod, 2013). Furthermore, large individual differences can be found in the frequency of phenomena like code-switching and code-mixing (Yip, 2013). All those aspects form potential variables influencing the speech acquisition and skills in bilingual children (Cruz-Ferreira, 2012; Goldstein & McLeod, 2012); especially when considering phonological acquisition from an emergentist perspective (cf. Section 1.1).

Some of these variables have been confirmed to impact on bilingual speech acquisition, while others have not or the findings were mixed. The AoA of the L2, for example, seems to have an effect on children's phonological performances at a certain age range only. Young bilingual children's phonological acquisition (i.e., up to age three/four) has been found to not be influenced by the AoA of L2 (Goldstein, Bunta, Lange, Rodriguez, & Burrows, 2010; Holm & Dodd, 2006), whereas an effect could be found in older children (i.e., from age four/five onwards; Morrow et al., 2014). This difference may be due to the different levels of language proficiency in L1 children had achieved by the time of acquisition of L2 (Iglesias & Rojas, 2012). However, regression analyses in Kim, Ballard and McCann (2017) revealed, that AoA of the L2 was not a predictor of 16 3;1- to 5;11-year-old Korean-English bilinguals' PCC-scores. Hence, the influence AoA has on bilingual children's phonological skills may be marginal. Further, research has shown that variability in language experience may affect linguistic output in a language-specific fashion (Gathercole & Hoff, 2007; Law & So, 2006).

In the following section four different factors that have frequently been discussed in the literature regarding their influence on bilingual children's phonological performance are outlined. First, gender is discussed as an impacting factor. Second and third, the influence of the amount of input and output in a bilingual's two languages are described. Finally, the role language proficiency may play in children's phonological competences is presented.

1.2.4.1 *The influence of gender on phonological acquisition*

The evidence for gender having an influence on speech accuracy (e.g., PCC) in typically and atypically developing monolingual children is mixed. Gruber (1999), for example, studied 24 American-English speaking children aged 3 - 5 years with speech delay and could not find any statistically significant gender differences in the children's PCC, PCC-A (PCC-Adjusted), or PCC-R (PCC-Revised) scores at any time-point. Studying a significantly larger sample ($N = 1{,}359$), Ege (2010) found typically developing Turkish-speaking 2;0 - 7;11-year-old boys to produce more errors on average than girls but the difference was not statistically significant. An investigation of 684 3;0- to 6;11-year-old British children's typical phonological development by Dodd et al. (2003) revealed an effect of gender in favour for girls on speech accuracy (PCC) and phonological acquisition (phoneme acquisition). However, this was only evident for the oldest age group (5;6 - 6;11-year-olds). Similarly, Schäfer and Fox (2006) identified gender differences in the consistency skills of 28 two-year-old German-speaking children. On average, boys in the 2;6 - 2;11-years age group were found to perform less well than their female peers. This difference became apparent as soon as children with an inconsistency rate of above 40% (i.e., those who are considered to show a deviant-inconsistent phonological disorder according to Dodd (1995)) were excluded from the analyses. Fox and Dodd (1999), in contrast could not find any statistically significant gender-related difference in 177 typically developing German-speaking children.

Although less frequently, gender effects on phonological skills have also been investigated for bilingual children. In a study on 40 Cantonese-English bilingual children Holm (1998) did not find gender to have a significant power on the children's PCC-scores in neither their English nor their Cantonese productions. Similarly, Prezas, Hodson and Schommer-Aikins (2014) could not find any gender-related differences in their 56 Spanish-English bilingual's percentages of occurrences of phonological deviations. However, Munro, Ball, Müller, Duckworth and Lyddy (2005) discovered gender-specific differences in the accuracy of consonant realisations in Welsh. Their Welsh-English bilingual girls were better at realising the lateral /s/ whereas boys performed better at realising /χ/. Thus, there seems to be a tendency for typically-developing boys and girls to differ on phonological measures but these differences have only seldom reached statistical significance. Male gender, however, was identified as a risk factor for SSD in monolinguals (Campbell et al., 2003; Eadie et al., 2014; Shriberg, Tomblin, & McSweeny, 1999; Wren, Roulstone, & Miller, 2012). These studies do not provide insight into gender-specific performances in children with SSD, though.

1.2.4.2 *The influence of language input and output patterns on phonological acquisition*

The language(s) children perceive through their environment and the pattern of this exposure (frequency, amount, quality etc.) are crucial variables for understanding their language development and linguistic output (Bhatia & Ritchie, 1999). The more direct contact (i.e., interaction with another speaker) a child has to a certain language the more they will learn it (Pearson, Fernandez, Lewedeg, & Oller, 1997). Per definition, bilingual children are exposed to and use both of their languages in their everyday lives (Chilla et al., 2010; Grosjean, 1998; Scharff Rethfeld, 2013) but the amount and quality of this input and output varies from one individual to another, across their languages and over time (Gildersleeve-Neumann et al., 2008; Gildersleeve-Neumann et al., 2009; Hammer, Lawrence, Rodriguez, Davison, & Miccio, 2011; Pearson et al., 1997; Scharff Rethfeld, 2013; Willard, Agache, Jäkel, Glück, & Leyendecker, 2014). Bilingual families differ regarding the number of family members communicating with the child in the respective languages, the contexts in which the languages are used, the degree of code-switching that occurs and the attitude towards learning and maintaining each of the two languages (Backus et al., 2010; Gathercole, 2014; Hammer & Rodriguez, 2012; Kohnert, 2013).

Further, the number of different interlocutors and settings in which children are exposed to and use their languages outside of the family as well as their access to a more formal language register (e.g., through books and audio drama) vary considerably across children (Kohnert, 2013). Depending on this and the language skills of their interlocutors, differences among children may also occur in the diversity of dialects and accents they are exposed to as well as the degree of grammatical complexity and differentiation of the vocabulary children perceive and use (Becker, 2010; Grosjean, 2013; Hammer & Rodriguez, 2012). As for monolingual children (see functionalist/emergentist approaches in Section 1.1), it is anticipated that a diversity of interlocutors, settings and registers increases the provision of a broad range of linguistic structures in the child's input which will shape their receptive and expressive language skills (Gathercole, 2014; Place & Hoff, 2011). Khattab (2006), for example, found similarities between a child's and their mother's realisation patterns of certain phonetic aspects (voicing lead). Thus, a child's realisations may resemble/be interpreted as developmental patterns or speech errors while in fact being related to the quality of and the variability in the input reflecting their awareness to specific speech characteristics in their environment (Khattab, 2006).

In addition to the quality of input, quantitative aspects, such as the frequency or (relative) amount of language input and output, have been investigated regarding their influence on a child's speech acquisition. While studies examining the unique effect of frequency of language output on bilingual children's phonological performances have not revealed statistically significant results (Goldstein et al., 2010; Goldstein et al., 2005), those focusing on either the effect of the amount of input or a combination of input and output often yielded significant findings. Parra, Hoff and Core (2011), for instance, found a language-specific effect of the amount of language exposure (input) on the non-word repetition (NWR) skills of 41 Spanish-English bilinguals aged 22 months. Ruiz-Felter, Cooperson, Bedore and Peña (2016) evidenced a positive within-language correlation between the amount of current language input-output and five- to year-year-old Spanish-English children's phonological skills. The larger the amount of input-output was in a language, the better children performed on accuracy measures for consonants and vowels in that language.

Despite some further support for a positive relationship between language experience (i.e., language input and output combined or language dominance) and phonological (Law & So, 2006; Lydea, Brebner, & McCormack, 2014; Paradis, 2001; So & Leung, 2006), grammatical (Gutiérrez-Clellen & Kreiter, 2003) and lexical outcomes (Becker, 2010; Parra et al., 2011; Pearson et al., 1997), this relationship was not necessarily found to be valid for all measured speech aspects over time. Munro et al. (2005), for example, reported a positive effect of language dominance (here defined as frequency of usage as measured by language background parental questionnaires) on the accuracy of a large number but not all Welsh and English consonants. Further, Morrow et al. (2014) found English use at home to correlate with only a few of the phonological outcome measures (i.e., accuracy of some consonants and percentage of occurrence of some phonological patterns) of English language learner (ELL) children and at different time-points. Since similar results were found for months of exposure to English, the authors assumed that a greater amount of input in English (at home) does not necessarily lead to more stability in children's phonological skills.

Differences in study results may be explainable by the different ways in which the amount or frequency of input and output were measured and classified. Across studies, language dominance, for example, is defined in various ways but always categorises children into two or three groups (i.e., dominant in language A, dominant in language B, and sometimes: equally dominant in language A and B). Therefore, Goldstein et al. (2010) argue that it might be best to place children on a continuum regarding their experiences with and their competencies in their respective languages (e.g., 10% input/output in language A and 90% input/output in language B). This would allow for analyses regarding the effect of frequency of input/output along the continuum rather than in two/three categories.

1.2.4.3 The influence of language proficiency on phonological acquisition

It is well-known from language acquisition studies in monolingual children that the individual language domains (i.e., phonology, semantics, grammar) interact with each other during development (Davis & Bedore, 2013; Smith, McGregor, & Demille, 2006; Stoel-Gammon, 2011; Vihman, 2014). This interaction is further implied for bilingual children by Kehoe (2011) who notes that there is an articulatory basis to phonological production from which both languages of a bilingual may benefit.

Goldstein et al. (2010) examined the effect of two different language proficiency measures on the phonological skills in Spanish-English bilingual children: (a) parent-rated language proficiency and (b) mean length of utterances for words (MLU_w; i.e., a direct language proficiency measure). Children's results from both language proficiency measures were found to have significantly influenced their phonological skills in both languages. The extent of the effects, however, differed for Spanish and English. In general, effect sizes were larger for Spanish than for English and the percentage of vowels correct (PVC) was only significantly influenced in Spanish but not in English. The authors referred this difference to the general structural differences in the phonologies of the two languages (Goldstein et al., 2010). Cooperson, Bedore and Peña (2013) could confirm significant and strong correlations between Spanish-English bilingual children's phonological scores (as measured by the number of correct consonants, vowels and clusters) and their morphosyntactic skills (as measured by a cloze task and a sentence repetition task). Overall, their study revealed language-specific as well as cross-linguistic effects of language proficiency. Despite within-language measures formed the primary predictor for children's phonological performances in both languages (e.g., morphosyntactic performances in Spanish predicted phonological performances in Spanish), between-language measures were found to be second and third predictors for children's phonological skills (e.g., Spanish morphosyntactic skills predicted phonological performance in English). In contrast, the relationships Cooperson et al. (2013) found between children's phonological performances and their semantic abilities (receptive and expressive vocabulary), were low. The authors explained this dissimilarity with morphosyntactic output being more reliant on a child's phonological abilities than knowledge and output of the lexicon are (Cooperson et al., 2013).

Moreover, language proficiency (as measured by a morphosyntax screening including a cloze task and a sentence repetition task in both languages) significantly correlated with Spanish-English bilingual children's phonological performances on NWR-tasks in both languages (Goldstein et al., 2010; Summers, Bohman, Gillam, Peña, & Bedore, 2010). Similar to Cooperson et al.'s (2013) study, the relationship between vocabulary skills and children's phonological performances was also not significant in this study which contradicts findings by Parra et al. (2011) who found significant and language-specific correlations between NWR and vocabulary skills in Spanish-English bilinguals. Vocabulary skills in these studies, however, were examined using different assessment tasks. Whereas Cooperson et al. (2013) and Summers et al. (2010) used direct receptive and expressive vocabulary measures including children's association and categorisation abilities, Parra et al. (2011) applied indirect measures in form of caregiver reports, such as the McArthur-Bates Communicative Developmental Inventories (Fenson et al., 1994). Thus, overall the effect of language proficiency on a bilingual's phonological acquisition may depend on the way it was defined and measured.

1.2.5 Summary – Phonological acquisition in bilingual children

In summary, the main features of typical speech acquisition in bilingual children discovered over the last decades are:

- Bilinguals may use two separate phonological systems which mutually influence each other from a very early age. However, complete consensus among researchers is still lacking.
- The rate of speech acquisition in bilinguals is dependent on the linguistic features under study and children's experience with the respective languages. Thus, it is likely to differ from that in monolinguals.
- Bilingual children's phonological development and abilities are characterised by distributed skills across languages, the occurrence of uni-/bidirectional CLT and qualitative differences in the type and frequency of occurrence of phonological patterns compared to monolinguals. At this, the respective language combination has a key influence on the linguistic outcome.
- Socio-linguistic variables vary across bilingual individuals forming a very heterogeneous population. Many of these variables (e.g., gender, amount of input and output per language, language proficiency) have been found to influence bilingual children's phonological performances whereby effects may differ across the involved languages.

Since the linguistic systems of a bilingual speaker were found to interact with each other (e.g., in form of CLT; Fabiano-Smith & Goldstein, 2010b), knowing the specific characteristics of their phonological systems is essential for investigating and understanding bilingual children's speech productions. Therefore, the next section introduces the phonological systems of the languages relevant for this research (i.e., German and Turkish), their regional variations as well as the accentual differences in the speech of the Turkish-German population in Germany.

1.3 Introduction to German and Turkish phonology

Phonological systems vary considerably from one language to the other, especially if languages originate from different language families. The German and Turkish phonological systems, which are firmly in focus of the present research, are distinct from one another in many respects. Thus, the following subsections review and introduce the phonological systems of both, including the characteristics of the varieties of German and Turkish spoken in different regions of Germany and Turkey as well as by Turkish-German bilingual adult speakers.

1.3.1 German phonology

The German language belongs to the Western Germanic language branch of the Indo-European language family and is spoken by approximately 92 million speakers world-wide (Crystal, 2010). It is the official language in Germany, Austria, Switzerland, Luxemburg, Liechtenstein as well as in South Tyrol (North-Italy), Eastern Belgium and Namibia. Further, it is an acknowledged minority language in Brazil, Italy, Kazakhstan, Kyrgyzstan, Romania, Russia, Slovakia, Czech Republic, Hungary and South Africa (Hawkins, 2011). Due to frequent migration approximately 200 million people across the globe speak German as their second language (Crystal, 2010).

1.3.1.1 Consonants

The German language consists of 22 consonants. They belong to the following sound classes: stops (/p, b, t, d, k, g/), nasals (/m, n, ŋ/), fricatives (/f, v, s, z, ʃ, ç, x, ʁ, h/), affricates (/p͜f, t͜s/), and approximants (/l, j/). In addition, the German phonological system includes the glottal stop /ʔ/ which functions as a compulsory phone before word-initial or stressed syllable-initial vowels (e.g., *Ofen* (oven) [ˈʔoːfən], *Karaoke* (karaoke) [kaʁaˈʔoːkə]). Its phonemic status, however, is controversial as its omission does not change a word's meaning and its occurrence is predictable (Fagan, 2009; Hall, 2011). Therefore, the glottal stop is not considered as a German phoneme in the current research.

The affricates /t͜ʃ, d͜ʒ/ and the voiced postalveolar fricative /ʒ/ are often ranked among German phonemes as well (Hall, 2011). Whereas the occurrence of /d͜ʒ/ and /ʒ/ is restricted to loan words such as *Garage* (garage) [gaˈʁaːʒə] and *Jeans* (jeans) [d͜ʒiːns], /t͜ʃ/ does occur in native words such as *deutsch* (German) [dɔɪt͜ʃ]. Its phonemic status as an affricate, however, is frequently debated (cf. Grassegger, 2004; Hall, 2011; Wiese, 2011). For comparison purposes with monolingual data by Fox and Dodd (1999) and due to the lack of a sufficient number of items including this sound in the German phonology assessment PLAKSS-II (Fox-Boyer, 2014c), /t͜ʃ/ will not be considered as a German affricate in this research.

The phonemic status of /ç/, /x/ (which may also be realised as /χ/), and /ŋ/ as well as the classification of /p͜f, t͜s/ as affricates are often debated as well. The occurrence of /ç, x/ is position dependent with /ç/ only appearing after front vowels as well as the sonorants /l, n, ʁ/ and /x/ only occurring after central and back vowels (Wiese, 2011). In addition to their predictable occurrence, /ç, x/ share two phonetic features (i.e., manner of articulation and [-voicing]) hence, some authors classify them as allophones of each other (Grassegger, 2004; Hall, 2011; Hawkins, 2011; Wiese, 2011). For comparison purposes with monolingual data by Fox and Dodd (1999), however, they will be considered as two phonemes here.

The debate about the phonemic status of /ŋ/ is rooted in the high restriction of its contextual occurrence in German words. It only occurs in syllable- or word-final positions (Hall, 2011). When contemplating the occurrence of the German nasals before stops, /m/ is present before stops of all three places of articulation (bilabial, alveolar and velar) whereas /n/ and /ŋ/ only appear before alveolar and velar stops respectively. As a consequence, /ŋ/ is regarded as an allophone of /n/ in the sound sequences of /ng/ and /nk/ which is generated by a nasal assimilation and a deletion of /g/ (Wiese, 2011). However, since /ŋ/ contrasts with other German phonemes such as *sing* (sing) [zɪŋ] vs. *Sinn* (sense/meaning) [zɪn]) it is treated as a separate phoneme in this research.

According to Hawkins (2011) there is no clear basis for the distinction of /p͜f, t͜s/ from other frequent consonant clusters as /ps, tʃ, ks/. However, as minimal pairs for /p͜f/ and /t͜s/ are evident in the German language, they are treated as affricates here.

Furthermore, three allophones of the voiced uvular fricative /ʁ/ can be found. In some regional dialects, it is produced as a trill (/r/ or /ʀ/) and vowelised as /ɐ/ when following a vowel in syllable- or word-final position (e.g., *Karte* (map) [ˈkaːɐtə], *Teller* (plate) [ˈtɛlɐ]) as well as when occurring as the first segment of a word-final consonant cluster (e.g., *Pferd* (horse) [p͜fɛːɐt], *Wurst* (sausage) [vʊɐst]) (Hall, 2011).

One characteristic process in the German language is the *Auslautverhärtung*, which means devoicing of all syllable- and word-final stops and fricatives (e.g., plural form *Tage* (days) [ˈtaːgə] but singular form: *Tag* (day) [taːk]) (Grassegger, 2004).

1.3.1.2 Vowels

The vowel system of German comprises 14 vowel phonemes /i, y, u, ɪ, ʏ, ʊ, e, ø, o, ɛ, œ, ə, ɔ, a/ and three diphthong phonemes /aʊ, aɪ, ɔɪ/. Additionally, diphthong-like vowel combinations can occur with the allophonic variation of /ʁ/ (i.e., [ɐ]) in words such as *ihr* (her) [iːɐ], *Uhr* (clock) [uːɐ], and *Ohr* (ear) [oːɐ]. The occurrence of the schwa /ə/ is mostly predictable as it never appears in stressed syllables. However, due to its contrastive properties as, for instance, in the words *genau* (exactly) [gəˈnaʊ] and *genial* (brilliant) [genˈjaːl] (Hall, 2011), it is considered as a phoneme in the present research.

In German, vowel length is a phonemic feature to differentiate meaning. The phonetic difference is a laxer articulation for the short vowels and a tenser articulation for the long vowels (Hawkins, 2011). Some examples for minimal pairs are *Teller* (plate) [ˈtɛlɐ] - *Täler* (valleys) [ˈtɛːlɐ], and *kann* (he/she/it can) [kan] - *Kahn* (barge) [kaːn].

1.3.1.3 Syllable structure and stress

German syllables are rather complex allowing a number of consonant clusters (CC) in the onset and coda ($C_{0\text{-}3}$ V $C_{0\text{-}4}$).
Onset clusters can consist of the following combinations:

- obstruent + liquid/nasal/obstruent
- fricative + stop + liquid

Coda clusters may include combinations of:

- liquid + obstruent
- liquid + liquid/nasal (+obstruent)
- obstruent + obstruent (+obstruent)
- nasal + obstruent
- liquid + nasal + /st/ (flexive morpheme)
- sonorant + nasal + /st/ or /t͜s/ (Fagan, 2009; Fox, 2004).

The German language is known for its large number of compound words which contribute to the frequent occurrence of multisyllabic words whose number of syllables is practically unlimited (Fagan, 2009).

Prosodic characteristics of the German language are a varying stress pattern including stress on final (e.g., *Elefant* (elephant) [eləˈfant]), penultimate (e.g., *Zitrone* (lemon) [t͜siˈtʁoːnə]) and antepenultimate syllables (e.g., *Schmetterling* (butterfly) [ˈʃmɛtɐlɪŋ]) but generally favouring the trochaic stress pattern (Wiese, 2011). In some rare cases the stress pattern can have a distinctive function as, for instance, in the word *umfahren*: (knock over) [ˈʊmfaːʁən] vs. (drive around) [ʊmˈfaːʁən] (Grassegger, 2004).

1.3.1.4 Accent variations in German consonants

There is a large number of variations to *Hochdeutsch* (High Standard German) spoken in Germany which vary from subtle accents to clear dialects such as *Bairisch* (Bavarian), *Schwäbisch* (Swabian), *Westfälisch* (Westphalian), to name just a few. *Hochdeutsch* is the language of education and has been adopted to a greater or lesser extent by many German speakers in their everyday conversations (Durrell, 1992). Usually there are still some features of regional variations perceivable in a speaker's pronunciation. For the present research, the phonetic-phonological changes on consonants in two areas of Germany are particular important: (a) the differences to *Hochdeutsch* in *Norddeutsch* (spoken in Northern Germany incl. the regions of Hamburg and Lübeck) and (b) the differences to *Hochdeutsch* in *Rheinisch* (spoken in the Rhineland incl. the region of Remscheid). The regions Northern Germany and Rhineland constitute the areas where data collection for this research took place, so that the variations to *Hochdeutsch* spoken there are briefly described in the following.

Phonetic-phonological differences to *Hochdeutsch* in *Norddeutsch*:

- Word-initial /p͜f/ is predominantly realised as [f] (e.g., *Pferd* (horse) /p͜fɛɐt/ → [fɛɐt]) (Barbour & Stevenson, 1990; Durrell, 1992).
- Word-final <g> may be realised as [k, ç, x] (e.g., *weg* (away) /vɛk/ → [vɛç]) (Barbour & Stevenson, 1990; Durrell, 1992).
- In word-final positions /ŋ/ may be produced as [ŋk] (e.g., Schmetterling (butterfly) /ˈʃmɛtɐlɪŋ/ → [ˈʃmɛtɐlɪŋk] (Durrell, 1992).
- Word-initial /t͜s/ may be reduced to [s] (e.g., *Zeitung* (newspaper) /ˈt͜saɪtʊŋ/ → [ˈsaɪtʊŋ]) (Barbour & Stevenson, 1990).
- The unstressed word-final syllables /ən/ and /əl/ may be reduced and assimilated (e.g., *Regen* (rain) /ˈʁeːgən/ → [ʁeːgŋ], *Apfel* (apple) /ap͜fəl/ → [ap͜fl]) (Durrell, 1992; Goltz & Walker, 1990).

Phonetic-phonological differences to *Hochdeutsch* in *Rheinisch*:

- The palatal fricative /ç/ is predominantly realised as [ʃ] (e.g., *Küche* (kitchen) /ˈkʏçə/ → [ˈkʏʃə]) (Barbour & Stevenson, 1990; Durrell, 1992; Keller, 1961; Newton, 1990).
- Word-initial /p͜f/ is predominantly realised as [f] (e.g., *Pferd* (horse) /p͜fɛɐt/ → [fɛɐt]) (Barbour & Stevenson, 1990; Durrell, 1992).

1.3.2 Turkish phonology

The Turkish language belongs to the Turkic group of the Altaic language family and is spoken by approximately 63 million people (Crystal, 2010). Turkish has official status in Turkey and Northern Cyprus and is the official language in regions of Romania, Greece, Macedonia and Bulgaria (Kornfilt, 2011). A large population of Turkish-speaking people lives in Western Europe (Göksel & Kerslake, 2005), and approximately 2.85 million people with a Turkish migration background reside in Germany (Statistisches Bundesamt, 2017). Precise numbers of speakers of Turkish in Germany do not exist but a broad estimate can be gained from the results of Chlosta, Ostermann and Schroeder's (2003) study in which 27.2% of their 5,213 multilingual participants attending primary schools in Essen[6] indicated they speak Turkish as an additional language to German.

1.3.2.1 Consonants

With 21 consonants, the Turkish phonology system contains a similar number of consonants as the German language. These can be allocated to the following sound classes: stops (/p, b, t, d, k, g/), nasals (/m, n/), fricatives (/f, v, s, z, ʃ, ʒ, ɣ, h/), affricates (/t͜ʃ, d͜ʒ/), approximants (/l, j/) and the tap (/ɾ/). Although the voiced post-alveolar fricative /ʒ/ only occurs in loan words and may be realised as /d͜ʒ/ in conversational speech (Ege, 2010; Lewis, 1967) it is considered to be part of the Turkish phoneme inventory by many authors (Kopkallı-Yavuz, 2010; Topbaş, 2007), and so in the present research. The existence of the voiced velar fricative /ɣ/ which is allocated to the Turkish letter <ğ> (also called 'yumuşak-g', i.e., 'soft-g') has caused several debates. Following acoustic studies, Kopkallı-Yavuz (2010) argues that it is not pronounced at all in Modern Turkish and only functions to lengthen the preceding vowel in words such as *ağaç* (tree) [ʌːt͜ʃ], *kurbağa* (frog) [kuɾˈbʌː], and y*ağmur* (rain) [jʌːˈmuɾ]. This is in line with previous reports from Lewis

[6] Essen is one of the large German cities with more than 500,000 inhabitants (Stadt Essen, Amt für Statistik, Stadtforschung und Wahlen, 2017). It is based in the federal state of North Rhine-Westphalia, which is populated by a large proportion of Turkish nationals (BAMF, 2016).

(1967) and Topbaş (2007). However, it does occur in some non-standard dialects as a voiced velar fricative (i.e., [ʌɣˈʌt͡ʃ], [kuɾbʌɣˈʌ], [jʌɣˈmuɾ]) (Brendemoen, 1998; Göksel & Kerslake, 2005; Topbaş & Yavaş, 2006). In the present research either pronunciation of the 'yumuşak-g' is scored as correct. However, it is not treated as a separate phoneme and thus its acquisition was not explored in the present research.

Many of the Turkish phonemes have frequently occurring allophones. The sounds /k, g/ are realised as their palatal counterparts [c, ɟ] when preceding or following front vowels but realised as [k, g] in all other sound environments (Kopkallı-Yavuz, 2010). However, in loan words they may also be realised as [c, ɟ] when neighboured by back vowels (Lewis, 1967). The voiced labio-dental fricative /v/ is mainly realised as [ʋ] in word-final and intervocalic position. The latter is especially common when the second vowel is rounded (Kopkallı-Yavuz, 2010; Topbaş, 2007). Göksel and Kerslake (2005) as well as Lewis (1967), however, describe the realisation of /v/ between vowels with the bilabial glide /w/. Since, there is a lack of phonetic studies to provide evidence for one or the other (Kopkallı-Yavuz, 2010) both allophonic variations are accepted in the present research. With front vowels /l/ is realised as a 'clear l', whereas it is velarised with back vowels as a /ɫ/ or a 'dark l' (Göksel & Kerslake, 2005; Lewis, 1967; Yavaş & Topbaş, 2004). Further, the alveloars /t, d, n/ are no true alveolars but produced dentally by trend (Lewis, 1967). The tap has three frequent allophones. In syllable-initial positions it is realised as a voiced fricated /ɾ/, word-finally it is usually realised as a voiceless fricated tap /ɾ̥/ whereas it is usually realised as a voiced alveolar tap /ɾ/ intervocalically (Lewis, 1967; Yavaş & Topbaş, 2004). All of these allophonic variations were accepted in the present research.

Some authors highlight that the glottal fricative /h/ may be realised as /ç/ in environments with front vowels, that is preceding or following a front vowel, and is often pronounced as /x/ in positions following a back vowel (Göksel & Kerslake, 2005; Lewis, 1967; Tekin, 2011). The /h/ is only pronounced as such in syllable-initial position preceding a back vowel. It is also frequently silent when occurring between a vowel and a consonant, where it leads to a lengthening of the preceding vowel (Göksel & Kerslake, 2005; Lewis, 1967). Further, the sound /j/ is barely heard as such when following a vowel and usually simply causes a diphthong-like sound (Lewis, 1967). Realisations as appropriate diphthongs are therefore accepted in the present research.

1.3.2.2 Vowels

The Turkish language consists of a significantly smaller number of vowel phonemes than the German language. The eight existing vowels, /i, y, ɛ, œ, ɯ, u, o, ʌ/, are all phonemically short, so in contrast to German vowel length is not used as a distinctive phonemic feature in Turkish (Lewis, 1967; Topbaş, 2007; Zimmer & Orgun, 1999). There are only three conditions in which long vowels occur:

a) in borrowed words (e.g., from Arabic)
b) vowels following <ğ> + consonant, front vowels following /j/, back vowels followed by <ğ>
c) to emphasise a word in speech (Lewis, 1967).

There is some debate about the height and backness properties of Turkish vowels in the literature so that different symbols of the International Phonetic Alphabet (IPA) have been allocated to some vowels (e.g., <a> → /a/ or /ʌ/, <ö> →/œ/ or /ø/) (Kopkallı-Yavuz, 2010). Darcy and Krüger (2012) argue that these different descriptions are an indication for the existence of a variety of allophones for Turkish vowels. However, Kopkallı-Yavuz (2010) emphasises that in previous research vowels were assessed in isolation or in varying phonetic environments which were not controlled for. To clarify the controversial findings, she conducted an acoustic analysis of Standard Turkish vowels for which the phonetic environment was controlled for and her classification of Turkish vowels (see beginning of this subsection) is referred to throughout this research.

A further, distinctive feature of Turkish phonology is the assimilatory process of vowel harmony which rules the formation of Turkish suffixes and words. Vowel harmony determines that all vowels within one domain, apart from the first syllable, have at least one feature in common (Hall, 2011). For Turkish, these features are backness and roundedness (Kopkallı-Yavuz, 2010). Thus, if a suffix is added to a word stem its vowel adapts in backness and roundedness to the vowel in the preceding syllable, such as *kitap-lar* (books) [citʌp-ˈlʌɾ], *cami-ler* (mosques) [d͜ʒʌ:mi-ˈlɛɾ]) (Göksel & Kerslake, 2005).

1.3.2.3 Syllable structure and stress

The Turkish syllable structure is less complex than the German. Consonant clusters hardly occur and if so are only present in the coda (C_{0-1} V C_{0-2}). Final consonant clusters in Turkish may consist of combinations of:

- sonorant + obstruent
- [-voice] fricative + stop
- /k/ + stop in loanwords.

In the rare case where consonant clusters occur in the onset this is in loan words. Usually a vowel is inserted in or before a cluster to split it up (e.g., *istasyon* (station) /istasˈjon/, *tiren* (train) /tiˈrɛn/) (Kornfilt, 2011; Lewis, 1967; Topbaş, 2007).

Due to Turkish being an agglutinative language it contains a large number of multisyllabic words with an unlimited number of syllables per word (Kornfilt, 2011).

In comparison to the varying stress pattern in German, Turkish is very regularly stressed on the final syllable, an exception being loan words and some names of cities (e.g., *İstanbul* [isˈtʌnbuɫ], *Ankara* [ˈʌnkʌɾʌ], *lokanta* (restaurant) [loˈkʌntʌ] (Kornfilt, 2011; Topbaş, 2007).

1.3.2.4 Accent variations in Turkish consonants

To what extent the linguistic system of Standard Turkish variates across different Turkish dialects is an underrepresented issue in the literature. Brendemoen (1998) names the low status of dialects in the Turkish society and the fact that large-scale surveys on dialectal variations have often been regarded as politically sensitive as the main reasons for them being limited. Nevertheless, some information on the phonological differences in dialects and Standard Turkish have been reported in textbooks on Turkish Grammar or mentioned in language/phonological acquisition studies. The variations indicated in these resources are summarised below:

- Velarisation of syllable- and word-final /h/ as [x] after back vowels (e.g., *siyah* (black) /siˈjʌh/ → [siˈjʌx]) (Lewis, 1967; Zimmer & Orgun, 1999).
- Frication of syllable- and word-final /k/ to [x] (e.g., *tavuk* (chicken) /tʌˈʊuk/ → [tʌˈʊux]), which occurs in most Anatolian dialects (Lewis, 1967).
- Occasionally velarisation of /n/ as [ŋ] as a co-articulatory effect when preceding velar consonants (e.g., *renk* (colour) /ɾɛnk/ → [ɾɛŋk]) (Kopkallı-Yavuz, 2010) but also in other positions (Brendemoen, 1998).
- (De-)Voicing of initial alveolar and velar stops when neighbouring back vowels occurs in some northeast dialects of Turkish (e.g., *karpuz* (watermelon) /kʌɾˈpuz/ → [gʌɾˈpuz]; *davul* (camel) /dʌˈʊuɫ/ → [tʌˈʊuɫ]) (Brendemoen, 1998).

1.3.3 Phonetic-phonological characteristics in the speech of adult Turkish-German bilingual speakers

Languages are flexible and underlie constant changes in all of their domains (Barbour & Stevenson, 1990). This can be observed especially in situations where language systems are in contact through multilingual speakers (Gagarina, 2014a; Martin, 2009; Muysken, 2013; Sankoff, 2004). The following two subsections outline the observed and hypothetical phonetic-phonological characteristics of German and Turkish by adult

Turkish-German bilingual speakers and will thus provide insight into the potential quality of children’s speech input in German and Turkish by their parents/relatives.

1.3.3.1 Variations of the German spoken by adult Turkish-heritage speakers in Germany

Most literature on structural changes to the phonetics and phonology of German induced by contact to the Turkish phonological system through Turkish-German bilingual speakers originated from the area of socio-linguistics, especially the study of ethnolects (Androutsopoulos, 2001; Bücker, 2007; Tekin & Colliander, 2010). Observations are generally rather anecdotal and lack empirical evidence from large-scale studies. Nevertheless, authors have agreed on the following variations from *Hochdeutsch* to occur in Turkish German (‘Türkendeutsch’), a variation of German often spoken by Turkish-heritage speakers in Germany, which affect consonants:

- Coronalisation of /ç/ to [ʃ] (e.g., *Licht* (light) /lɪçt/ → [lɪʃt]) (Androutsopoulos, 2001; Bücker, 2007). This variation is also evident in regional variations of *Hochdeutsch* (especially, in the Rhineland) (Barbour & Stevenson, 1990; Keller, 1961). However, since it has also been found to occur in Turkish German spoken in Berlin it is also recognised as part of this ethnolect (Androutsopoulos, 2001).
- Deaffrication and voicing of /t͜s/ to [z] (e.g., *Zebra* (zebra) /t͜seːbʁa/ → [zeːbʁa]) (Androutsopoulos, 2001).
- No realisation of glottal stops (e.g., *Ofen* (oven) /ʔoːfən/ → [oːfən]) (Androutsopoulos, 2001).
- Apical realisation of /ʁ/ in certain syllable positions, predominantly in codas (e.g., *Winter* (winter) /vɪntɐ/ → [vɪntɛr]) (Androutsopoulos, 2001; Tekin, 2011).
- Biphonemic realisation of /ŋ/ as [ng] (Tekin, 2011).
- Vowel epenthesis in German consonant clusters as Turkish does not allow initial consonant clusters (Androutsopoulos, 2001; Tekin, 2011).

These characteristics of adult speaker’s pronunciation of German provide valuable information on the types of phonetic-phonological differences to *Hochdeutsch* that are to be expected in Turkish-German bilingual children’s realisation of German as adults function as their phonetic-phonological role models.

1.3.3.2 Variations of the Turkish spoken in Germany

Due to the socio-linguistic history of Turkish German a large variability in Turkish-heritage speakers’ registers and inventories has developed which make generalisations to this population very difficult (Küppers, Şimşek, & Schroeder, 2015). Despite some extensive research on the syntax and morphology (Backus et al., 2010; Pfaff, 1993; Pfaff, 1991), lexicon and literacy (Schroeder & Şimşek, 2010) of the Turkish spoken in Germany, there is a lack of studies on the phonological changes that are likely to occur (Haig & Braun, 1999; Küppers et al., 2015). Since Turkish-German bilingual adults with varying ages of acquisition of German were found to not necessarily achieve accent free speech in one or both of their languages (Stangen, Kupisch, Proietti Ergün, & Zielke, 2015), phonological differences are very likely. Some first evidence comes from studies on the intonation patterns of young Turkish-German speakers. These were found to use intonation patterns in both Turkish and German that are typical for neither language (Queen, 2001). However, since the exact phonological changes to the Turkish spoken in Germany still need to be explored, contrastive analyses of the German and Turkish linguistic systems are currently the only possible way to derive hypotheses about phonological changes. Some of these hypothetical variations are listed below[7]:

- The German phonological process of *Auslautverhärtung* may have an influence on a speaker's realisations of word-final /v/ and /z/ in Turkish, e.g., *muz* (banana) /muz/ → [mus], *ev* (house) /ɛʋ/ → [ɛf].
- The use of the allophonic variations of /k, g, l, v/ may be reduced as these do not occur in German, e.g., *köpek* (dog) /cœˈpɛc/ → [kœˈpɛk], *tavuk* (chicken) /tʌˈʋuk/ → [tʌˈvuk].
- The German sound /ŋ/ and its use in positions preceding velar stops on syllable boundaries or in CC may be transferred to Turkish, e.g., *renk* (colour) /rɛnc/ → [rɛŋk].

1.4 Phonological acquisition in German and Turkish

As outlined in Section 1.1, the phonological acquisition in children shows some cross-linguistic similarities (e.g., acquisition of simple aspects before complex aspects) but also evidences significant influence from the respective ambient language. This may become apparent in quantitative as well as qualitative aspects. The following subsections introduce the most recent findings on the phonological acquisition in monolingual German- and Turkish-speaking children as well as the current state of knowledge on the phonological acquisition in Turkish-German bilingual children.

[7] This list is not considered to be exhaustive.

1.4.1 Phonological acquisition in monolingual German-speaking children

The most recent data on the phonological acquisition in typically developing German-speaking children is provided by cross-sectional studies by Fox-Boyer and colleagues. The authors tested around 800 German-speaking children aged 1;6 - 5;11 years between the years of 1999 and 2014. Based on single-word naming and consistency assessments the studies provide an overview of the phonetic and phonemic acquisition of consonants (Fox & Dodd, 1999), the phonemic acquisition of consonant clusters (Fox-Boyer, 2014b; Fox & Dodd, 1999; Schaefer & Fox-Boyer, 2016), the acquisition of consistency of word production (Schäfer & Fox, 2006), the use and suppression of developmental phonological patterns (Fox-Boyer, 2014b; Fox-Boyer & Schäfer, 2015) as well as the stimulability of isolated phones (Kubaschk, Fox-Boyer, & Klann, 2015). The publications by Fox-Boyer (2014b), Fox-Boyer and Schäfer (2015), and Schaefer and Fox-Boyer (2016) include data from several previous studies that were reanalysed and supplemented by new data. The other listed studies present original analyses of separate data.

Fox and Dodd (1999) reported that the 177 German-speaking children in their study had acquired (75% of the children in an age group produced the sound phonemically correct) and mastered (90% of the children in an age group produced the sound phonemically correct) the majority of German consonants by the age of 3;0 (see Table 1.3). Some fricatives and affricates were acquired/mastered slightly later with all German consonants being acquired by the age of 4;0 and mastered by the age of 5;0 years the latest.

Table 1.3: Consonant (phoneme) acquisition in German-speaking children (adapted from Fox and Dodd, 1999)

Age group	75%-criterion	90%-criterion
1;6 - 1;11	p, b, t, d, m, n,	p, d, m
2;0- 2;5	v, s, z, h	b, n
2;6- 2;11	k, g, ŋ, f, x, ʁ, l, j, p͜f	t, k, ŋ, f, v, s, z, x, h, l
3;0 - 3;5	ç, t͜s	g, ʁ, j, p͜f
3;6 - 3;11	ʃ	t͜s
4;0 - 4;5		ç
4;6 - 4;11		ʃ

If compared cross-linguistically (cf. McLeod, 2007), the acquisition of the German consonants /d, z, v/ is interestingly early. For example, in many languages the phoneme /z/ is not mastered (90%-criterion) before the age range of 3;6 - 4;0 years (or even later) (Ben-David & Berman, 2007; Dyson & Amayreh, 2007; Khattab, 2007; Mennen &

Okalidou, 2007; Ota & Ueda, 2007; Rahilly, 2007; So, 2007; Zajdó, 2007). Fox and Dodd (1999) explain this phenomenon by the concept of phonological saliency (So & Dodd, 1995; Zhu & Dodd, 2000b). The consonants /d/ and /v/ may be more salient for German-speaking children as they very frequently occur in young children's early vocabulary and their parents' speech (e.g., in words such as *da* (there) /da/, *wer, was, wo* (who, what, where) /vɛɐ, vas, voː/, *wauwau* (early onomatopoeia for dog) /ˈvaʊvaʊ/). In contrast, the sound /z/ may be more salient as its occurrence in German is distributive with /s/ (Fox, 2004). These examples highlight the role of the ambient language's phonology on children's phonological acquisition (cf. Section 1.1).

According to data from Fox-Boyer (2014b), the majority of German initial consonant clusters is acquired by the age of 3;6 and mastered by the age of 4;0 (see Table 1.4). As data from Schäfer and Fox-Boyer (2016) on the consonant cluster acquisition of two-year-olds demonstrates, children as young as 2;0 - 2;5 years are already able to produce a large number of German clusters. Further, their findings highlight the early acquisition of three-element consonant clusters (i.e., at the same time as two-element clusters) if fronting and backing errors of /ʃ/ (which also occurred for singletons) were accepted as correct.

Table 1.4: Acquisition of initial consonant clusters in German (adapted from Fox-Boyer, 2014b)

Age group	75%-criterion	90%-criterion
2;6 - 2;11	bʁ, bl, kl, gl	kl
3;0 - 3;5	tʁ, dʁ, kʁ, gʁ, kn, kv, fʁ, fl, ʃm, ʃn, ʃv, ʃʁ, ʃl, ʃp, ʃt	tʁ, dʁ, kv, gl, fʁ, fl, ʃm, ʃn, ʃv, ʃʁ
3;6 - 3;11	ʃpʁ, ʃtʁ	bʁ, bl, gʁ, ʃp, ʃt, ʃl
4;0 - 4;5		kn, kʁ, ʃpʁ, ʃtʁ

Detailed acquisition data on final consonant clusters do not exist but one cross-sectional (Fox & Dodd, 1999) and one longitudinal study (Lleó & Prinz, 1996) have investigated children's cluster reduction patterns. Up to the age of 2;5 years, final clusters were usually reduced to the first element. After this age, children only rarely reduced final consonant clusters (Fox, 2016).

Regarding the phonological patterns present in children's speech, Fox-Boyer and Schäfer (2015) reported eight different pattern types to occur in the speech of typically developing children aged 2;6 - 2;11 years. The most frequent patterns, which can also be found in the speech of significantly older children, are fronting of /ʃ/ followed by consonant cluster reduction and backing of /ʃ/ to [ç]. An overview of the phonological patterns and their age of suppression is presented in Table 1.5.

Table 1.5: Phonological patterns in typically developing German-speaking children (adapted from Fox-Boyer, 2014b and Fox-Boyer and Schäfer, 2015)

Phonological pattern	**Age groups**						
	2;0 - 2;5	2;6 - 2;11	3;0 - 3;5	3;6 - 3;11	4;0 - 4;5	4;6 - 4;11	5;0 - 5;5
Fronting /ʃ/ → [s]	■	■	■	■	■	■	
Weak syllable deletion[1]	■	■	■	■	■	■	
Fronting /ç/ → [s]	■	■	■	■			
CC reduction	■	■	■	■			
Contact assimilation[2]	■	■	■	■			
Fronting /k, g/ → [t, d]	■	■	■				
Assimilation	■	■					
Backing /ʃ/ → [ç]	■	■					
Final cons. deletion	■						
SF consonant deletion	■						
Glottal substitution	■						
Deaffrication	■						
Fronting /ŋ/ → [n]	■						
Stopping (Fric./Affr.)[3]	■						

Note: **Cons.**: consonant, **SF**: syllable-final, **Fric./Affr.**: Fricative/Affricate, **[1]**mainly affected weak syllables in pre-stressed positions (Fox-Boyer, 2014b), **[2]**Fox-Boyer (2014b) notes that the presence of contact assimilation in typical speech may be an artefact of the items in the test rather than a phonological pattern typically present in children's speech acquisition, **[3]**This pattern occurred infrequently and was often indistinguishable from the pattern of assimilation (e.g., /p͜fɛɐt/ → [tɛɐt], Fox-Boyer and Schäfer, 2015), pattern is present in at least 10% of this age group

Prosodically, children have most difficulties with the acquisition of weak syllables in pre-stressed positions (Fox-Boyer, 2014b). This stress pattern deviates from the otherwise very prominent trochee in German and is thus prone to cause some issues during development. However, as soon as children have acquired this syllable structure, it can be assumed that they have mastered all other stress patterns present in German as well (Fox-Boyer & Schäfer, 2015). Fox-Boyer and Schäfer (2015) further note that from an age as young as 2;6 years, children only rarely demonstrate a phonological pattern in all possible occurrences (i.e., to 100%). For most patterns, occurrences range from 10 - 90% of all possible occurrences in the speech sample. For these data, a pattern was considered to be present in children's speech if it occurred at least five times in the single-word naming sample. Fox (2016) additionally reported children's average number of phonological pattern tokens that occurred less than five times in the test (i.e., only 1 - 4 times; referred to as *infrequent variants (InfVar)* in this research). Results showed that these significantly decrease with children's age with 22.8 tokens on average at 2;6 - 2;11 years to 7.4 tokens on average at 3;6 - 3;11 years. This illustrates the pace with which German-speaking children acquire the phonological system of their native language.

1.4.2 Phonological acquisition in monolingual Turkish-speaking children

The most recent studies on the phonological acquisition in Turkish-speaking children cover the age range of 1;0 - 8;0 years and include 2,889 participants from cross-sectional and 104 participants from longitudinal investigations in total. The aspects under study encompassed the phonetic and phonemic acquisition of consonants (Ege, 2010; Topbaş, 1997; Topbaş & Yavaş, 2006; Yalcinkaya, Muluk, & Budak, 2010; Yavaş & Topbaş, 2004), the acquisition and reduction patterns of final consonant clusters (Topbaş, 2007; Topbaş & Kopkalli-Yavuz, 2008), the percentage of consonants correct (Topbaş & Yavaş, 2006), the production of phonological patterns (Topbaş, 1997, 2006; Topbaş & KopkallI-Yavuz, 1997; Topbaş & Yavaş, 2006) as well as NWR-skills (Topbas, Kacar-Kutukcu, & Kopkalli-Yavuz, 2014). The publications of Topbaş and Yavaş (2006), Topbaş (2006, 2007) and Topbaş and Kopkallı-Yavuz (2008) refer to the same data pool whereas the other listed publications refer to separate data.

As Table 1.6 illustrates, Turkish-speaking children acquire and master most of the phonemes of their languages by the age of 3;5. Some later acquired consonants of German (e.g., /ʃ/) are acquired comparably early in Turkish-speaking children.

Table 1.6: Consonant (phoneme) acquisition in Turkish-speaking children (adapted from Topbaş and Yavaş, 2006)

Age group	75%-criterion	90%-criterion
1;6 - 1;11	p, b, t, d, k, m, n, j,	b, t, d, m, n, j
2;0 - 2;5	g, v, s, z, ʃ, l, t͜ʃ, d͜ʒ	p, k, g, v, l, t͜ʃ, d͜ʒ
2;6 - 2;11	f, ʒ	s, ʃ,
3;0 - 3;5	z, h, ɾ	f, z, ʒ, h
3;6 - 3;11		ɾ

Note: Topbaş and Yavaş (2006) reported the phoneme acquisition per syllable position. In order to keep the data comparable to those for German, Topbaş and Yavaş's table was simplified. A consonant is presented as acquired (75%)/mastered (90%) if it was reported as respective for at least two different syllable positions.

However, the ages of acquisition vary slightly depending on the assessment items used, the scoring method applied and the sample size assessed. Ege (2010) and Yalcinkaya et al. (2010), both having assessed larger cohorts, reported some consonants to be later acquired. For example, Ege (2010) stated, based on data of 1,359 participants, that children master the sound /ɾ/ between the ages of 4;6 and 6;0 years depending on the position in the word (i.e., word-final position was acquired first and syllable-initial position last) and the gender of the child (girls acquired the sound a little earlier than boys). Generally, she found Turkish-speaking children to acquire most consonants word-finally first compared to other positions in the word. She explained this phenomenon by the highly predictable stress pattern (nearly exclusively word-final stress) as well as the reduced number of consonants allowed in word-final position (i.e., no voiced stops and affricates) in Turkish which she relates to the concept of phonological saliency (Zhu & Dodd, 2000b).

Turkish phonotactics only allow syllable-final consonant clusters which generally occur very rarely. Thus, it is not surprising that Turkish-speaking children acquire these structures relatively late compared to German-speaking children (see Table 1.7). Further, Yavaş and Topbaş's (2004) study revealed that the first element of the cluster is more prone to omissions or distortions during children's acquisition. This is in contrast to the cluster acquisition pattern of German-speaking children (see Section 1.4.1).

Table 1.7: Acquisition of final consonant clusters in Turkish (adapted from Topbaş, 2007)

Age group	Coda CC
3;0 - 3;6	nt͜ʃ, nk
3;6 - 4;0	ft, st, lp, lk
4;0 - 4;6	
4;6 - 5;0	ɾt, ɾk

The phonological patterns observed in typically developing Turkish-speaking children show some similarities with those found in German-speaking monolinguals (e.g., the presence of fronting of velars and postalveolars) but also provide evidence for language-specific patterns, as can be seen from Table 1.8. Participants in Topbaş and Yavaş's (2006) study were further reported to have overcome the majority of phonological patterns on average by 3;6 years, with the reduction of final consonant clusters as well as liquid deviations persisting a bit longer. A few patterns reported for the typical phonological acquisition in Turkish are considered to be atypical for German-speaking children (e.g., metathesis, affrication). Data on the average number of pattern tokens in children of a certain age group have not been reported yet.

Table 1.8: Phonological patterns in typically developing Turkish-speaking children (adapted from Topbaş and Yavaş, 2006)

Phonological pattern	Age					
	1;6 - 1;11	2;0 - 2;5	2;6 - 2;11	3;0 - 3;5	3;6 - 3;11	4;0 - 4;5
Final CC reduction	■	■	■	■	■	
Liquid deviation	■	■	■	■	■	
Stopping (Fric./Affr.)	■	■	■	■		
Syllable deletion[1]	■	■	■	■		
Consonant deletion[2]	■	■	■			
Assimilation	■	■	■			
Fronting postalveolars & velars	■	■	■			
Word-final devoicing /v, z/	■	■	■			
Deaffrication	■	■				
Affrication	■	■				
Voicing	■	■				
Metathesis	■	■				
Backing alveolars	■	■				
Reduplication	■					

Note: **Fric./Affr.**: Fricative/Affricate, **[1]**stressed and unstressed syllables, **[2]**syllable-initial and syllable-final, pattern is present in at least 10% of this age group

1.4.3 Phonological acquisition in Turkish-German bilingual children

To date, two pilot studies with small sample sizes (Salgert et al., 2012; Ünsal & Fox, 2002) and one larger-scale project (Naş, 2015) have cross-sectionally investigated the speech acquisition in typically developing Turkish-German bilingual children. Naş (2015) reports data on the acquisition of Turkish consonants only, having assessed 136 typically developing Turkish-German bilingual children aged between 3;9 and 6;3 years. According to his data, the fricatives /s, z, ʃ, ʒ/, the affricates /t͜ʃ, d͜ʒ/ and the tap /ɾ/ were those consonants that were not acquired by any of the included age groups, whereas the following consonants seemed to be acquired by most age groups: /m, n, p, b, t, d, k, g, f, h, j/. However, it cannot be gathered from the author's presentation whether children's phone or phoneme acquisition was examined and what cut-off criteria for the acquisition of a consonant were applied. Therefore, the findings have to be considered with caution.

Salgert et al. (2012) and Ünsal and Fox (2002) provided data on children's phonological acquisition in both languages including phone and phoneme inventories as well as types of phonetic-phonological patterns. In both studies, the performances of Turkish-German bilingual children were compared to existing data on monolinguals. Although targeting slightly different age ranges (i.e., 2;10 - 3;6 and 3;10 - 4;6 years in Salgert et al. and 4;11 - 6;1 years in Ünsal and Fox) than Naş's (2015) study both pilot studies reported similar consonants to be phonetically and phonemically late acquired in Turkish. In general, the pilot studies provided evidence for a partly delayed phonological acquisition when compared to monolingual peers in both Turkish and German. Mainly affected by this deceleration were fricatives and affricates; sound classes that are considered to be rather complex (Jakobson, 1968) and cross-linguistically often reported to be later acquired (Fabiano-Smith & Goldstein, 2010a; Goldstein & Washington, 2001; Morrow et al., 2014; Stow & Pert, 2006). By the age of five, however, children's phonetic and phonemic inventories were almost complete in both Turkish and German (Ünsal & Fox, 2002). The remaining phones in German were /s, z, ç/ and /ts/, whereas only /ç/ was also not acquired phonemically yet. In Turkish, the fricatives /s, z, ʒ/ were still missing to the 90% criterion in the phone inventories, with only /ʒ/ not being acquired phonemically yet (Ünsal & Fox, 2002). Further, these two studies revealed that some phonological patterns were suppressed at a later age in bilinguals than reported for monolinguals. Participants in Salgert et al.'s (2012) study, for example, still fronted postalveolar fricatives in Turkish, a phonological pattern typically overcome by the age of 2;6 in Turkish monolinguals. Similarly, Salgert et al.'s older participants (i.e., 3;10 - 4;6 years) as well as Ünsal and Fox's children (4;11 - 6;1 years) still reduced initial consonant clusters in German although German monolinguals usually overcome this phonological pattern by the age of 4;0. In general, all phonological patterns the children produced could also be found in monolingual data of each language. An exception was CLT which was evident for consonants, vowels, phonological patterns as well as stress patterns (Salgert et al., 2012). Furthermore, Salgert et al. (2012) reported an exceptionally large number of vowel distortions in both languages. In German, this mainly affected the Schwa (/ə/), which was predominantly realised as a full vowel (i.e., /ɛ/ or /e/). In Turkish, this usually affected vowels which were perceptually similar to German counterparts (e.g., the Turkish /ɯ/ was realised as the German [ɪ]).

The limited data available illustrate the current lack of reliable normative data on Turkish-German bilingual phonological acquisition. Methodologically sound studies on large cohorts as well as longitudinal investigations are still needed to validate these initial findings and to gain a better understanding of the developmental pattern in the speech acquisition of Turkish-German bilingual children. This is what the current research project aimed to address.

2 LITERATURE REVIEW II – Differentiation between typical and atypical phonological acquisition

This chapter focuses on evidence-based procedures to assess bilingual children's speech abilities and to differentiate typical from atypical phonological development.

Due to their heterogeneous composition, the population of bilingual children presents the examiner with a number of challenges when it comes to assessing and evaluating their phonological skills. The first of the following two subsections elaborates on the general considerations for assessing children's speech, the gold standards for constructing phonology assessments as well as specific international recommendations for evaluating bilingual children's speech. The section concludes with an overview of the existing assessment tools for examining phonological skills in Turkish-German bilingual children. The second subsection is dedicated to the identification of speech sound difficulties (SSD) in children. It begins with an introduction to clinical markers used with monolinguals and the hitherto identification of markers for SSD in bilinguals. Subsequently, the diagnostic relevance of longitudinal investigations as well as of different assessment tasks are described and discussed. The section concludes with the aims, questions and hypotheses of this research.

2.1 Assessment of phonological skills

Over the last 20 years, assessment procedures frequently used with monolingual children have been discussed regarding their suitability for bilingual children (Thordardottir, Rothenberg, Rivard, & Naves, 2006; Yavaş & Goldstein, 1998). Despite the basic assessment principles being the same for children acquiring one or more languages – i.e., to describe the child's phonological system and to decide whether it deviates from typical development and requires professional support (Wyatt, 2012; Yavaş & Goldstein, 1998) – several additional aspects need to be considered for the phonological assessment with bilinguals. These as well as general test construction criteria for phonology assessments are presented in the following.

2.1.1 General test construction principles for phonology assessments

When designing high quality ability assessments, internationally developed requirements regarding psychometric properties (i.e., validity, normative sample and reliability) should be met and the concept of the assessment needs to be theoretically grounded. In this subsection the different aspects of validity, considerations for normative samples as well as the facets of reliability applying to phonology assessments are presented.

2.1.1.1 Validity

A test is considered to be valid if it actually measures what it intends to measure in an accurate way (McCauley & Swisher, 1984). To fulfil this criterion, test developers should ensure that the assessment is designed to meet the different aspects of validity, such as *content relevance and coverage, predictive and diagnostic utility, interpretive meaningfulness* including convergent coherence, discriminant distinctiveness, and task generalisability (AERA, APA, & NCME, 1999; Messick, 1980). The different facets of validity are viewed to be part of the comprehensive concept of 'construct validity' in this research rather than separate 'forms of validity' as construct validity *"[…] integrates criterion and content considerations into a common framework for testing rational hypotheses about theoretically relevant relationships*" (Messick, 1980, p. 1015).

Regarding the need to meet *content relevance and coverage* requirements, researchers have pointed out a significant number of linguistic criteria that should be considered and met when designing a phonological assessment tool for (monolingual) children. These may vary slightly depending on the type of analyses to be applied to the collected data, including compiling consonant inventories, PCC-analysis, and phonological pattern analysis (e.g., Eisenberg & Hitchcock, 2010; Kirk & Vigeland, 2015) as well as in the way the speech sample is collected, for example, via confrontation naming or connected speech sample (Morrison & Shriberg, 1992). If the assessment is also intended to be used in the clinical decision-making process, a number of additional aspects need to be considered: (a) evidence for reliability; (b) availability of normative basis; and (c) diagnostic accuracy as, for example, sensitivity and specificity (Eisenberg & Hitchcock, 2010). Thus, the purpose of the assessment needs to be clearly defined (Friberg, 2010).

In order to ensure task generalisability test developers should choose an assessment task which represents the ability or behaviour under study in an unbiased way (Messick, 1980). The appropriateness of single-word naming has been debated with regard to its representation of a child's true speech abilities. Since it only refers to the production of isolated words, it has been argued that the behaviour observed with this elicitation method may not be sufficient to fully reflect abilities in natural speech output (Klein & Liu-Shea, 2009; Morrison & Shriberg, 1992; Stoel-Gammon & Dunn, 1985). However, the consistent and structured way in which the examiner obtains speech data with single-word naming tests as well as the possibility to allow for norm-referenced scores may make well-designed tests equally representative of a child's speech skills as spontaneous connected speech productions (Bernhardt & Holdgrafer, 2001; Masterson, Bernhardt, & Hofheinz, 2005; Wolk & Meisler, 1998). Within the scope of this thesis only those guidelines referring to single-word naming tasks are presented and discussed.

One way of ensuring that the assessment is representative of the children's speech abilities is to ascertain that it allows for a large enough number of occurrences of phonological features for subsequent analyses to be comprehensive and accurate. For this, Stoel-Gammon and Dunn (1985) recommend that the assessment encompasses a minimum of approximately 75-100 words when used in addition to a continuous speech sample. Ideally, these items include each phoneme of the respective language in as many positions in the word as phonotactically possible, including clusters, to procure a full picture of children's production skills of individual phonemes (Bernhardt & Holdgrafer, 2001; Stoel-Gammon & Dunn, 1985; Yavaş & Goldstein, 1998). This is especially important as the accuracy of speech sound realisation has been found to be influenced by the sound's position in the word as well as by its phonetic and phonemic environment (Bankson, Bernthal, & Flipsen, 2013; Edwards & Beckman, 2008; Stoel-Gammon & Dunn, 1985). Further, children's accuracy in realising consonants was found to be influenced by whether they occur in a stressed or unstressed syllable (Bernhardt & Holdgrafer, 2001; Kirk & Demuth, 2006) so that speech sounds are ideally assessed in words with varying stress patterns (Edwards & Beckman, 2008; Stackhouse & Wells, 1997). Morphophonemic contexts, however, should be avoided as sounds in these contexts have been found to be less accurately produced than those in singleton contexts and may impede a clear distinction of morphological and phonological impact (McLeod, 2012b; Shriberg & Kwiatkowsky, 1994; Stackhouse & Wells, 1997).

Regarding the sampling frequency of each phoneme, occurrences of at least one to five times (in different words) have been recommended in order to allow for a sufficient evaluation basis (McLeod, 2012b; Yavaş & Goldstein, 1998). Further, several studies by James and her colleagues (2016; 2008a; 2001b; 2008b) demonstrated that multisyllabic words have been especially discriminative in the evaluation of children's speech skills, thus words of three or more syllables are recommended to be included in a phonological assessment.

Beyond focusing on single segments, the possible occurrence of phonological patterns should be taken into consideration as children typically exhibit variations from the adult target form that apply to more than just a single phoneme, i.e., classes of speech sounds (Kirk & Vigeland, 2014). In order to provide a representative picture of the child's phonological system, opportunities for production of each of the phonological patterns common in typical and atypical phonological development should be included, with four to five opportunities widely recommended (Dodd, 2005; Kirk & Vigeland, 2015;

McReynolds & Elbert, 1981; Stoel-Gammon & Dunn, 1985). The determination of this cut-off, however, is arbitrary and varies slightly from one publication to another. According to Kirk and Vigeland (2014), a test should further provide an equal number of opportunities for each error pattern to occur to ensure validity regarding its internal structure. This, however, is a significant challenge as phoneme frequency varies considerably within languages (Hawkins, 2011; Kopkallı-Yavuz, 2010). An imbalance in the number of possible occurrences of phonological patterns will therefore be inevitable if all other linguistic criteria were to be met.

In general, the items of an assessment need to be selected carefully and in relation to the purpose of the test as well as the analyses intended with it as, for example, a determination of the phonetic inventory of a child requires a slightly different density of certain phonological features than a phonological pattern analysis (Eisenberg & Hitchcock, 2010; Flipsen & Ogiela, 2015; Kirk & Vigeland, 2015).

To ensure good content validity words should be selected that are familiar to children of the age range under study since young children's phonological abilities are linked to vocabulary knowledge (Stoel-Gammon & Williams, 2013) and potentially co-occurring language difficulties in the children could bias the results (Bernhardt & Holdgrafer, 2001; Flipsen & Ogiela, 2015; McLeod, 2012b; Stackhouse & Wells, 1997). Related to this is an analysis of the items' relative difficulty, both linguistically (Friberg, 2010) and in terms of familiarity (James, 2001b). To facilitate spontaneous naming, items of a speech assessment should be presented very clearly and in child-friendly form such as through coloured photos or ClipArts (McLeod, 2012b), as black and white line drawings have been found to be less effective for eliciting target items in young children (Bernthal, Grossman, & Goll, 1989). To further rule out the effect of tiredness and a decline in children's attention span the overall duration of the speech assessment should be kept to an appropriate level for the target age group (McCauley & Swisher, 1984).

Concurrent utility refers to the agreement between scores on a new test and a previously designed and valid assessment of the same kind, i.e., one that assesses the same construct (AERA et al., 1999). Correlations between these assessments need to be sufficiently high to demonstrate that assessments would generate similar results. However, if these correlations are very high and the new test does not show any advantages, for example, regarding time efficiency, over the one it is compared to, its usefulness and necessity is debateable (Anastasi, 1982). The issue with this facet of validity is that it cannot be applied to assessments which are the first of their kind or when no other linguistically or psychometrically suitable assessment tool is available for comparison, as the main motive to devise a new phonology assessment is the lack of well-designed instruments.

In contrast, *diagnostic or identification accuracy,* which refers to a test's sensitivity and specificity as well as its positive and negative predictive values, could be determined by the degree of agreement between expert judgements and a test's diagnostic categorisation of a child's performance (Kirk & Vigeland, 2014; McCauley & Swisher, 1984). In other words, diagnostic accuracy refers to the test's ability to reliably filter out children with SSD (sensitivity) and to clearly discriminate these from children whose phonological development progresses normally (specificity) (Lalkhen & McCluskey, 2008). Due to the importance of accurately identifying children with atypical development because of the significant social consequences for wrongly identified children, an ideal assessment should have a very low number of false-negative and false-positive cases (Friberg, 2010; Plante & Vance, 1994; Spaulding, Plante, & Farinella, 2006). This aspect of validity sheds light on the overall precision of diagnoses made with the test and is therefore considered to be the most important aspect, especially if the assessment was used for clinical decision-making (Friberg, 2010; Plante & Vance, 1994).

Further aspects to consider concern the test's ability to predict future performance on the same or related skills (*predictive utility)* (McCauley & Swisher, 1984) as well as whether the test displays *developmental trends* in a sample of children of different age groups (Anastasi, 1982). This information is especially interesting for clinical settings where comparisons to peers need to be made and response to intervention as well as future performances on, for example, reading abilities are to be evaluated (Flipsen & Ogiela, 2015). Although, older children would be expected to outperform younger children on a phonological assessment (at least up to a certain age), it needs to be borne in mind that many of a child's abilities and behaviours follow a developmental pattern (Kirk & Vigeland, 2014) so that this aspect of validity is considered to be relatively weak.

2.1.1.2 Normative sample

Normative samples provide the basis for comparisons of obtained data and are used as a reference in clinical decision-making. Therefore, it is necessary that the composition of participants in the sample is as representative of the variability in the general population under study as possible (Kirk & Vigeland, 2014). This means, if an assessment is designed for monolingual British English speakers and aimed to be used to differentiate between typical and atypical speech development in young children, it should include British English-speaking children of the particular age range (e.g., 2;0 – 7;11 years for speech development) and both genders who come from various socioeconomic backgrounds and a variety of regions in the UK where British English is spoken in order to be demographically representative (Flipsen & Ogiela, 2015; Friberg, 2010; Plante & Vance, 1994; Wyatt, 2012). Whether children with SSD should also be included in the sample so that comparative scores for both typically and atypically developing children were available and a possible sample truncation could be avoided (AERA et al., 1999; Kirk & Vigeland, 2014; McCauley & Swisher, 1984) or whether this would impede the diagnostic accuracy of the test (Peña, Spaulding, & Plante, 2006) is still a subject of debate. Nevertheless, sufficient information on the sample's composition need to be reported in the test's manual to allow for correct interpretations of the normative scores (Friberg, 2010).

Regarding the size of the normative sample, Flipsen and Ogiela (2015) emphasise that the overall size is less important than the number of participants in the respective subgroups that are used for clinical comparisons (e.g., age group, gender, typical and atypical children). Amongst others, McCauley and Swisher (1984) recommend that each subgroup should ideally encompass a minimum of 100 subjects to ensure a sufficiently large number for statistical analyses but also to represent the wider population's variability and distribution of scores on the test. To further provide test users with an estimate of the size of the variation shown they suggest that mean and standard deviations are reported for each subgroup included in the sample (McCauley & Swisher, 1984).

Normative samples need to be updated frequently as language and demographics change over time (Kirk & Vigeland, 2014). Rathvon (2004), for example, recommends an interval of approximately 12 years after which normative data need to be renewed.

2.1.1.3 Reliability

Test reliability relates to the consistency with which a test produces scores for a certain behaviour or ability across items (*internal consistency*), over time (*test-retest reliability*) and across different examiners (*inter-rater reliability*) (Anastasi, 1982; McCauley & Swisher, 1984). Since many social, environmental and administration aspects can have an influence on a test's outcome test designers need to ensure that those influences are kept to a minimum.

To determine a test's *internal consistency,* the scores its items produce are correlated with each other in a systematic way by either dividing the test in two equal parts and correlating a child's scores on both parts with one another (split-half-reliability) or by applying the coefficient alpha (Rust & Golombok, 1999). High correlation coefficients confirm that different items reliably measure the same general construct (Anastasi, 1982). This way of measuring *internal consistency* is especially useful for tests that produce a numerical score per item or rather similar maximum scores per item. However, in the case of a phonology assessment, words included in the assessment are purposely chosen to vary in terms of their phonemic and phonotactic structure to allow for the occurrence of a variety of phonological patterns so that maximum scores per item cannot be predicted. Therefore, this aspect of reliability is considered to be less suitable for single-word naming assessments which are aimed to measure the child's phonological abilities.

Test-retest-reliability considers the impact a child's constitution on a specific day (e.g., attention span, tiredness, mood) may have on their performance in the test (McCauley & Swisher, 1984). Test designers should investigate the effects of these parameters by administering the same assessment, in the same way with the same client repeatedly within a short interval. For instruments that are intended to assess developmental skills an interval of a maximum of two weeks is to be aimed for so that both learning and maturational biasing effects can be avoided (Sánchez, 2006). High correlations between the total scores but also on an item-by-item comparison indicate a good stability of the test over time (McCauley & Swisher, 1984).

Good *inter-rater reliability* ensures that the test is very robust (i.e., does not produce significantly different scores) against both the examiner's interactional style, behaviour or rapport to the child as well as their interpretation of the child's performance (Flipsen & Ogiela, 2015). This aspect of reliability is either assessed by having different examiners administering the test with the same children within a short interval or by letting different examiners analyse the performances of the same children (Kirk & Vigeland, 2014). A high correlation coefficient or point-by-point agreement (i.e., .90) indicates that test results only vary to a minimum if the test is administered or analysed by various examiners (McCauley & Swisher, 1984). This reliability procedure can also be applied to the same examiner to ensure *intra-rater reliability* (McLeod, 2012b).

In addition to the forms of reliability outlined above, Flipsen and Ogiela (2015) citing Hutchinson (1996) point out the importance of reporting a test's standard error of measurement (SEM). This measure refers to the standard deviation of scores a child would produce if he/she was asked to repeat the test several times. Since phonology tests are not only used to assess a child's speech abilities one single time but also to document and monitor their developmental progress with and without intervention, SEM values provide a useful guidance for practitioners to interpret test results.

In order to achieve high test-retest and inter- as well as intra-rater reliabilities it is essential to provide clear instructions for examiners so that they can administer the test in a consistent and standardised way (McCauley & Swisher, 1984). Further, it is helpful to provide sufficient information on the qualifications and training needed to professionally administer the test in the manual (McCauley & Swisher, 1984). The latter is particularly important as some tests do not only require a certain degree of familiarity with their instruction but also with the specific forms of analyses.

AERA et al. (1999) further recommend to report reliability results for each subgroup included in the normative sample.

2.1.2 Considerations for the assessment and analysis of bilingual children's speech

Due to the involvement of more than one language and the heterogeneity of the bilingual population a number of additional aspects need to be taken into account when assessing and analysing bilingual children's speech skills. This section outlines the challenges involved, discusses the use of assessments for monolinguals with bilingual populations and presents international guidelines for the construction of phonology assessments for bilingual children.

2.1.2.1 Challenges in the assessment of bilingual children

Many researchers, international associations and expert panels have argued for assessing bilingual children in all of their languages in order to gain a full picture of their overall speech abilities (e.g., De Lamo White & Jin, 2011; Fredman, 2011; Kohnert, 2013; McLeod, 2012a; RCSLT, 2007; Scharff Rethfeld, 2005; IEPMCS, 2012; Thordardottir et al., 2006; Yavaş & Goldstein, 1998). This is particularly important because: (a) the phonological systems of languages differ – as can be seen from the outline of German and Turkish phonology in Section 1.3; (b) bilingual children's phonological skills are usually different in both of their languages, often demonstrating an uneven ability across languages (Holm et al., 1999); (c) cross-linguistic associations might occur in the children's speech (Fabiano-Smith & Barlow, 2010); (d) children's exposure to, use of, and proficiency in their languages are likely to vary (Kohnert, 2010); and (e) children's acquisition of languages may have started at different points in time, thus, increasing the likelihood of acquisition differences (McLeod et al., 2017).

One of the main challenges with assessing children's speech across languages is that SLTs are rarely competent speakers of their bilingual clients' first language. In a survey by Lüke and Ritterfeld (2011), German SLTs indicated that the multilingual children they were seeing (2,951 children in total) spoke 49 different languages in addition to German. Although this variety does not affect a single clinic it gives an idea of the diversity of (first) languages spoken by the multilingual clients SLTs treat in Germany. Since communication barriers may impede the administration of speech assessments in the child's respective languages, including the analysis and evaluation of their performances (Thordardottir et al., 2006) and prevent SLTs from gathering crucial information on the client's case history (e.g., Holm & Dodd, 1999a; Jordaan, 2008; Mennen & Stansfield, 2006; Williams & McLeod, 2012), trained support personnel such as interpreters, translators etc. who have knowledge about the child's languages and cultures would be helpful to assist the SLT (Isaac, 2007; Kohnert, 2013; Langdon & Quintanar-Sarellana, 2003; Wyatt, 2012).

Besides these personnel factors, there is a dearth of assessments and normative data on the typical speech acquisition in many languages and dialects. Whereas some languages have been studied extensively regarding monolinguals' acquisition of speech sounds (e.g., English, Spanish, German etc.), others are not as well-researched, for example, Polish (Dodd, Holm, & Li, 1997; McLeod et al., 2013; Wyatt, 2012; Yavaş & Goldstein, 1998). Data on the typical speech acquisition in bilinguals is even more limited and speech assessments specifically designed for this population are the exception. Due to the lack of a comparative basis, little is known about the specific markers for speech delay and disorder in monolinguals who speak under-researched languages and in bilinguals in general (Williams & McLeod, 2012).

2.1.2.2 Appropriateness of using monolingual phonology assessments with bilingual children

On grounds of a lack of speech assessments specifically designed for bilingual populations, monolingual assessments in each language are generally used to examine the children's speech competences (Kritikos, 2003; Skahan et al., 2007; Thordardottir et al., 2006). The question then arises as to whether phonological assessments designed for monolingual children can be considered appropriate for the use with bilingual children. From a theoretical point of view, the application of monolingual assessments can be justified by the facts that bilingual children have been reported to have separate phonological systems from as early as the babbling stage (Maneva & Genesee, 2002) and that their pattern of speech and language acquisition (i.e., developmental milestones) is broadly in line with that of monolingual children – with the exception of a few bilingual-specific patterns (Fabiano-Smith & Barlow, 2010; Goldstein et al., 2005; Holm et al., 1999; Lin & Johnson, 2010). From a practical point of view, the advantages are that monolingual speech assessments exist for a larger variety of languages than assessments for bilingual children (McLeod, 2012a), they are generally easily accessible and SLTs are familiar with their administration and analysis – at least if those assessments are in the SLT's first language. In many cases, monolingual assessments are also well-designed linguistically and have been standardised so that psychometric criteria are often at least basically met (Flipsen & Ogiela, 2015; Friberg, 2010; Kirk & Vigeland, 2014).

The disadvantage of using monolingual speech assessments with bilingual children, however, is their tendency toward content and linguistic bias as well as their standardisation on monolingual norms. Content bias may occur for the following reasons: First, bilingual children have a different linguistic experience in both of their languages than their monolingual peers (Caesar & Kohler, 2007) since they are not two monolinguals in one person (Grosjean, 1989). Their lexical knowledge, for instance, is distributed over their languages encompassing words from only one of their languages as well as translation equivalents (Pearson, Fernandez, & Oller, 1993; Peña, Bedore, & Zlatic-Giunta, 2002). Thus, words to be known by monolingual children in a certain language may not necessarily also be known by their bilingual peers. Items in monolingual speech assessments, however, are usually selected to be representative of monolingual speakers' vocabulary so that the possible unfamiliarity with these items may have an impact on bilingual children's phonetic-phonological realisation of the words (McLeod, 2012a).

Second, the pictures used in monolingual speech assessments may not trigger the exact same words in monolingual and bilingual children due to their exposure to different cultural environments (Sánchez, 2006). An English tea cup, for example, may elicit the word 'tea' in a monolingual English child but potentially the word 'coffee' in a Turkish-English bilingual child due to the different types of pots that are used to pour tea and coffee in in Turkey and in England. Finally, the activity of naming pictures and interpreting them correctly is cultural-specific. Lidz and Peña (1996), for example, reported that Latino American children were more used to explain an object's function rather than naming it directly as this was the practice their mothers applied. Thus, items that are represented with very cultural-specific pictures and are not selected for the lexical knowledge of bilingual children may require further cues from the SLT (e.g., a prompt for repetition) to be elicited (Salgert, 2011).

Linguistic bias may occur if the child is only tested in one of their languages or if the dialect or accent of the assessing professional does not reflect the child's experience with this language (Kohnert, 2013).

2.1.2.3 Guidelines for the assessment setting and the design of phonology tests for bilinguals

Further aspects to consider before administering a speech assessment with a bilingual child are of sociocultural nature and regard the choice of assessment tools and delivery service. Yavaş and Goldstein (1998) advocate the use of culturally sensitive material to enhance the children's performance and a culturally sensitive behaviour towards the client and their family members to improve parental cooperation (Battle, 2012; Scharff Rethfeld, 2013). In addition, the SLT should acquire an in-depth knowledge about the phonologies of the child's languages in order to: (a) know which phonological features in particular need to be addressed; (b) correctly identify possibly occurring cross-linguistic interactions (transfer); and (c) confidently differentiate these interactions from atypical development (Fabiano-Smith & Goldstein, 2010b; Fabiano & Goldstein, 2005; Preston & Seki, 2011; Yavaş & Goldstein, 1998). If the phonological system of language A, for example, only allows devoiced consonants in the coda whereas in language B both voiced and devoiced consonants are possible, the child might devoice all consonants in the coda independent of the language he/she speaks. This would result in atypical speech output in language B and might be misdiagnosed as a phonological disorder. Given the knowledge about both languages' phonology, however, this can simply be accounted for by a transfer of a phonotactic restriction from language A to B.

Similarly, dialectal variations need to be accounted for during the whole diagnostic process (Goldstein & Fabiano, 2007, February 13; Goldstein & McLeod, 2012; Yavaş & Goldstein, 1998). Not considering these aspects would result in misleading diagnoses as dialect influenced speech may lead to a later acquisition of consonants, lower PCC-scores, a higher percentage in the occurrence of phonological patterns as well as a later overcome of phonological patterns (Phoon, Abdullah, & Maclagan, 2012). Additionally, examiners need to familiarise themselves with the potential phonetic-phonological and lexical changes that occur in the minority language due to being in contact with the majority language (Backus et al., 2010; Chilla et al., 2010). The Turkish spoken in Germany, for instance, varies from the high standard Turkish spoken in Turkey on all linguistic levels (Backus, 2006 and cf. Section 1.1.2.5). This would need to form the standard for comparison used in the assessment.

Monolingual data are a rather inappropriate source for reference as they represent a relatively homogeneous population regarding culture, AoA and input patterns (De Lamo White & Jin, 2011). Bilingual children usually achieve lower scores on monolingual assessments and may demonstrate different phonological patterns which may lead to misinterpretations (De Lamo White & Jin, 2011; Dodd et al., 1997; Thordardottir et al., 2006; Yavaş & Goldstein, 1998). Hence, in order to clearly differentiate between delays and disorders, normative data of bilinguals with the same language combination is the ideal prerequisite (Holm et al., 1999). Further, examiners need to find a way to deal with code-switching during the assessments as this may especially occur when children are very young and/or have to name unfamiliar items (McLeod, 2012a).

When the international recommendations for constructing assessments (McLeod et al., 2013) and the (dis)advantages of the use of monolingual assessments with bilingual children are taken into account a newly designed phonology assessment for bilingual children would need to fulfil the following criteria if it was to be used to assess one of the child's languages:

- The assessment task used should be culturally sensitive for the respective child. So, if picture-book reading or rather single-word naming is an unusual task but story telling is more common in the child's culture and language, eliciting connected speech samples should be preferred over a naming task (Yavaş & Goldstein, 1998).
- It should be ensured that all items are representative for the child's culture and vocabulary and are displayed with culturally sensitive illustrations to prevent erroneous naming and avoidable elicitation cues (McLeod, 2012a). In order to provide those, consultations of speakers of the respective language such as family members, bilingual SLT-colleagues or interpreters are recommended (Bernhardt & Holdgrafer, 2001; McLeod, 2012a).

For the evaluation of the child's performances two further linguistic aspects need to be carefully considered to avoid misinterpretations: (a) cross-linguistic interactions (transfer) need to be identified by the examiner (Yavaş, 1998) and (b) the specific interactions between the majority and minority language, which frequently occur due to languages being in contact, need to be evaluated regarding phonetic and phonological differences to the standard versions of the languages (Backus, 2006).

2.1.3 Existing phonology assessment tools for Turkish-German bilingual children

As the previous section has emphasised, assessment tools specifically designed for bilingual children are to be preferred over assessments for monolingual children of each language. To date, three phonology assessment tools have been published for the population of Turkish-heritage children living in German-speaking countries. These comprise two screenings and a phonology test which all target children's phonological skills in Turkish only. These are briefly introduced and their appropriateness for diagnosing SSD in the broader Turkish-German bilingual population is discussed in the following sections. Appendix A includes a comprehensive table of all aspects considered for the evaluation of the assessments' suitability.

2.1.3.1 Wiener Lautprüfverfahren für Türkisch sprechende Kinder (WIELAU-T)

The WIELAU-T [Viennese sound evaluation tool for Turkish-speaking children] (Lammer & Kalmár, 2004) is a screening tool for Turkish-speaking (pre-)school children in single-word naming format, which comprises a set of 34 items (nouns). The items evaluate all Turkish phonemes but the [ʒ] and one consonant cluster. Each phoneme is assessed at least twice and in different positions of the word. The consonant cluster is assessed once in syllable-final position. One- to four-syllabic words are included.

All items are displayed through photos of the objects. To support the German-speaking examiner in the administration of the assessment the target words are written orthographically and in IPA on the back of the pictures. On an additional audio CD, the items' target pronunciations can be listened to. Recommended analyses are segment orientated. The WIELAU-T is neither theory-bound nor standardised and does not include norms. It is based on the clinical knowledge and experience of the authors (Kalmár, 2006).

2.1.3.2 Screening der Erstsprachfähigkeit bei Migrantenkindern: Russisch-Deutsch, Türkisch-Deutsch (SCREEMIK 2)

The SCREEMIK 2 [Screening of first language skills in migrant children: Russian-German, Turkish-German] (Wagner, 2008) is a computerised speech and language screening designed and normed for 4;0- to 5;11-year-old Turkish-German and Russian-German bilingual children (Wagner, 2010). A subtask in this screening tool focuses on the assessment of speech sounds and includes a single-word naming and a sound discrimination task. To facilitate the evaluation of the bilingual child's speech skills by a monolingual German-speaking SLT, only shared sounds between Russian/Turkish and German as well as those who are of therapeutic relevance are included (Wagner, 2010). Further, all questions the computer "asks" the child in their native language are translated for the German-speaking SLT and IPA-transcriptions of the target response are provided.

The single-word naming task for Turkish consists of 36 items and covers 15 out of 21 different consonant phonemes in word-initial, word-medial and word-final position (if phonotactic restrictions allowed it). One- to three-syllabic words are included which predominantly consist of CV-syllable structures. These include a sound which only exists in some German dialects, the /r/. Children's realisations do not need to be transcribed as each item is only evaluated regarding the correct or incorrect pronunciation of a single target phoneme. This evaluation is conducted on-line and automatically summarised for each sound by the computer programme. The Turkish version has been normed with 388 Turkish-German bilingual children and its objectivity (i.e., test administration, analysis procedures, inter-rater reliability regarding analyses and interpretation), reliability (i.e., internal consistency, explorative factor analysis) and validity (i.e., criterion validity [agreement with nursery nurses' judgements, gender-related differences] and construct validity [age-related differences, inter-correlations of subtasks]) has been examined and approved (Wagner, 2008, 2010). Most of these analyses, however, refer to the whole screening (i.e., including the vocabulary and phoneme discrimination task) and are not presented for the single-word naming task separately. Further, criterion and construct validity were measured based on segment-orientated results and do not consider children's realisations of the whole word.

2.1.3.3 Türkisch-Artikulations-Test (TAT)

The TAT [Turkish articulation test] (Naş, 2010) is a single-word naming test designed to examine all Turkish speech sounds and one Turkish consonant cluster within a set of 118 items. The consonants are tested at least twice and mostly in different positions of the word. Each Turkish vowel occurs at least three times in the item list and the consonant cluster once. One- to four-syllable words are included.

So far, children between 3;0 and 14;3 years have been tested with this assessment. According to Naş (2010) all items are part of the vocabulary of Turkish-speaking children although references for this are not provided. Naş further states that the assessment can easily be administered by testers with no knowledge of Turkish as all items are transcribed in IPA and an additional audio CD provides the correct pronunciation of the items in standard Turkish and Turkish with a German accent.

The aim of the assessment is to identify SSD in Turkish-speaking children living in a German-speaking environment. According to Naş (2015), the test has recently been norm-referenced and validated for 136 typically developing and 56 atypically developing Turkish-German bilingual children aged 3;9 to 6;3 years. However, due to lack of clear methodological descriptions and explanations these findings are unable to be replicated and at least partly questionable. Naş (2015), for example, does not mention the type of analyses he has conducted to investigate the TAT's validity, fails to provide evidence for the appropriateness of the selected vocabulary and the types of analyses conducted to evaluate children's speech abilities are non-transparent.

The published assessment tools outlined above can help to provide a broad understanding of Turkish-German-speaking children's phonetic and phonological abilities in Turkish. The items include culturally sensitive words and are mainly presented with cultural-sensitive pictures and drawings. However, all of them lack several valuable criteria for a phonological assessment that follows international test construction guidelines as well as criteria that are particularly recommended for a multilingual assessment setting (see Appendix A). Firstly, the assessments do not meet all linguistic criteria. Whereas screening tools are expected to not examine all phonological aspects, test tools should do so. None of the presented assessments examines all Turkish speech sounds and consonant clusters to the recommended frequency (i.e., at least one to five times in the item set; McLeod, 2012b; Yavaş & Goldstein, 1998). Additionally, not all common phonological patterns known for Turkish monolingual children could occur at least five times. Thus, a reliable phonological pattern analysis according to Holm and Dodd (2006) could not be carried out.

Secondly, although the authors provide some information about how they selected the items based on their phonological features they only briefly comment on how they made sure that the items were familiar to Turkish-speaking children of the respective age groups living in German-speaking countries. Thus, all three instruments miss some important international requirements for phonology assessment tools and therefore do not seem to be appropriate for investigating the typical phonological acquisition in Turkish-German bilingual children and differentiating it from atypical development. A new phonology test for this population would have to fill these gaps and particularly focus on the linguistic criteria.

2.1.4 Summary – Assessment of phonological skills

This section has highlighted that one of the major challenges SLTs are facing when assessing and evaluating bilingual children's speech performances is the lack of suitable assessment tools. Phonology assessments for (Turkish-German) bilingual children do not only need to meet the necessary phonological criteria to ensure content relevance and coverage requirements but also require a careful selection of the included vocabulary to achieve a high item familiarity and to be culturally sensitive. Further, assessments should have been analysed on psychometric variables to ensure their validity, reliability as well as diagnostic accuracy with the target population and need to include normative data for reference.

A review of the strengths and weaknesses of all currently available Turkish assessment tools for Turkish-German bilingual children revealed the necessity for devising a new assessment tool if the phonological acquisition in this population were aimed to be evaluated validly and reliably. Therefore, a new Turkish phonology assessment which aimed to avoid the gaps of the existing tools and to fulfil the internationally recommended criteria (see Section 2.1.1) as far as possible was developed within this research project.

However, to ensure a high diagnostic accuracy, phonology assessments further need to be designed in a way that they allow for the detection of clinical markers for atypical phonological development in a given population. The current state of knowledge of clinical markers for identifying SSD in monolingual and bilingual children is presented in the next section.

2.2 Identification of speech sound difficulties in children

Information on the prevalence of SSD in monolingual children vary internationally from 6.4% (Broomfield & Dodd, 2004) to 15.3% (Campbell et al., 2003) depending on the age group under study and the cut-off values used (Law, Boyle, Harris, Harkness, & Nye, 2000). Given the fact that SSD occur across languages, and supposing that bilingualism itself does not entail SSD, similar prevalence rates for bilinguals should be assumed (Winter, 2001). However, a clear differentiation from atypical development is impeded since the variability in performances of typically developing bilinguals is considerably large. Considering the social, emotional and educational consequences of SSD, such as learning to read and write, school success, job prospects, peer-to-peer communications, child-parent relationships, and frustration when communication breaks down (e.g., Lewis et al., 2000; McCormack, McLeod, Harrison, & McAllister, 2010; McCormack et al., 2009; Nathan et al., 2004b; Preston, Hull, & Edwards, 2013), valid and reliable identification of children with suspected SSD is crucial. Especially, the question of what type of marker to use to determine a child’s performance as typical, delayed or disordered has attracted researchers’ interest (e.g., Holm & Dodd, 1999a; Holm et al., 1999). For monolingual children, several clinical markers for SSD have been identified (see below: Dodd, 1995; Shriberg, 1994; Shriberg et al., 2010; Stackhouse & Wells, 1997). The evidence base for the specific symptomatology of and markers for SSD in bilinguals, however, is still small and predominantly includes single case studies (Hambly, Wren, McLeod, & Roulstone, 2013).

2.2.1 Clinical markers for SSD in monolingual preschool children

Before specifically reporting on the clinical markers used to identify monolinguals with SSD, the three major classification approaches and frameworks described in the literature which have formed the basis for the identification of markers are introduced.

SSD of unknown origin have been described within the following major classification approaches: (a) the aetiological approach to which the Speech Disorders Classification System belongs (Shriberg, 1993, 1994; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997b; Shriberg et al., 2010; Shriberg & Kwiatkowsky, 1982a), (b) a psycholinguistically-grounded symptomatologic approach to which the model of differential diagnosis (Dodd, 1995, 2005) can be allocated, and (c) the psycholinguistic approach which includes the speech processing model from Stackhouse and Wells (1997).

Shriberg et al.'s (2010) system is based on genetics research and on the hypothesis that each subgroup of speech sound disorder can be allocated to a certain genetic phenotype. The Speech Disorders Classification System (SDCS) differentiates between eight subgroups of SSD which are classified by causal relationships between the following markers: (a) children's case history (i.e., aetiological factors) as well as their genetic predisposition for SSD (e.g., SLI, family history of SLI or SSD) and (b) their speech production performances assessed with the Madison Speech Assessment Protocol (MSAP), a specifically compiled list of tests and tasks including spontaneous speech samples and elicited single-word productions (Shriberg et al., 2010).

One issue with this classification system is that children with similar aetiologies may exhibit different speech processing difficulties so that an allocation to the subgroups does not necessarily allow for conclusions about the type and severity level of a SSD (Waring & Knight, 2013). Additionally, its clinical applicability is questionable as an allocation to the different subgroups is achieved via computer-generated algorithms within the Competence, Precision, and Stability Analysis (CPSA) framework (Shriberg et al., 2010; Vick et al., 2014). Fox, Dodd and Howard (2002) presented data that challenged Shriberg's (1994) aetiological classification system – the predecessor of the 2010 version of the SDCS – to be applicable to German-speaking children highlighting especially that (a) some children had more than one risk factor for SSD which made the allocation to a specific subgroup difficult, (b) the classification system did not take into account further risk factors such as sucking habits and pre- and perinatal factors, and c) that there was no empirical support that children with distinct aetiological factors would require different intervention approaches (Fox et al., 2002). Moreover, Shriberg and Kwiatkowsky's (1982b) severity rating for SSD and their extension from (1997a) was challenged cross-linguistically by data by Clausen and Fox-Boyer (submitted). The authors reported that the PCC-A-data of Danish-speaking children were not sufficient to categorise children into different subgroups of SSD as typically developing, delayed and disordered children (i.e., those with a consistent phonological disorder) demonstrated an overlap in their scores. Qualitative information from a phonological pattern analysis were necessary for a clear differentiation of the groups.

Up-to-date, the 2010 version of the SDCS is conceptualised for and has only been applied to monolingual English-speaking children (Shriberg et al., 2010). To use this classification system with bilingual children would require the collection of normative data on the relevant speech skills from the bilingual population under study since comparisons of monolingual and bilingual speech performances have revealed significant differences (e.g., Bunta et al., 2009; Fabiano-Smith & Goldstein, 2010a; Gildersleeve-Neumann et al., 2008; Goldstein & Washington, 2001). This includes the adaptation and development of assessment tools and tasks for the targeted bilingual population. Further, the application may be complicated by the parents' often limited conversational skills in the environmental language and/or by cultural differences which prevent them from providing the relevant aetiological and genetic information accurately (Battle, 2012; Mennen & Stansfield, 2006; Qualls, 2012).

In comparison, Dodd's (1995) classification system is based on the description of linguistic surface patterns (phonological patterns) which are used as clinical markers to categorise the children's performances into five different subgroups of SSD: *articulation disorder, phonological delay, consistent phonological disorder, inconsistent phonological disorder* and *childhood apraxia of speech*. For phonological disorders, Dodd (1995) differentiates between age-appropriate, delayed and atypical (deviant) phonological patterns. Her classification allows for an evaluation of children's speech performances that is independent of their clinical-genetic aetiology. Dodd, Leahy and Hambly (1989) further identified a specific speech processing deficit underlying each of the five subgroups and found out that children in these groups responded differently to therapy approaches (Crosbie, Holm, & Dodd, 2005). Thus, Dodd's (1995) classification system links linguistic surface patterns to psycholinguistic deficits which enables the examiner to formulate hypotheses about the type of disorder and allows for generating therapy goals. However, in order to identify those markers and to classify children's performances, assessment tools which allow for phonological pattern analyses as well as include a 25-words inconsistency test are required alongside normative data. It needs to be criticised, though, that the inconsistency test advocated by Dodd (1995) has not been validated for any language yet (Waring & Knight, 2013) and the stability of the five subgroups has not yet been systematically investigated over time (Rvachew & Brosseau-Lapré, 2012).

Following the assumption that the underlying deficits of SSD are language-independent (Dodd et al., 1997), Dodd's classification system has been proven to be valid cross-linguistically (e.g., Broomfield & Dodd, 2004; Fox & Dodd, 2001; So & Dodd, 1995; So & Dodd, 1994; Zhu, 2000) and has further been applied with bilingual children (Grech & Dodd, 2008; Holm & Dodd, 1999a, 1999b; Holm, Dodd, & Ozanne, 1997; Holm et al., 1999). This nevertheless required the availability of appropriate assessment tools to examine children's skills in the involved languages. Further, the existence of normative data on the respective populations under study was necessary to identify and discriminate children with SSD from those who were typically developing.

Representatives of the psycholinguistic approach, in contrast, do not intend to classify children's performances into several subgroups of SSD but to highlight children's individual strengths and weaknesses in the speech processing chain to determine the level where their processing breaks down (Stackhouse & Wells, 1993). The speech processing model a variety of speech input and output tasks are administered to test different hypotheses and profile the children's individual abilities and weaknesses. This requires an extensive assessment battery to assess the different speech processing levels as well as the child's phonological patterns (Stackhouse & Wells, 1997). The level of difficulty of the individual tasks in the assessment battery need to be adjusted to the child's chronological age to achieve reliable and valid results (Stackhouse, Vance, Pascoe, & Wells, 2007). Using the psycholinguistic speech processing model (Stackhouse & Wells, 1997) thus offers a unique possibility to examine children's skills on different speech processing levels and their relation to children's phonetic-phonological output. Hence, it is suited for identifying difficulties in the speech processing chain that may cause atypical phonological output but which are not captured with a single-word naming task alone. It thus allows the examiner to derive individual intervention goals for each child (Stackhouse & Wells, 1997). However, as Waring and Knight (2013) point out, this approach makes it difficult to predict the course of speech development and intervention success due to its exclusive focus on the individual.

Stackhouse and Well's (1997) speech processing model has been developed for and predominantly applied to monolingual English-speaking children although Fricke and Schäfer (2008) also used this model as the basis for their assessment of phonological awareness skills in monolingual German-speaking children. Although Stackhouse and Wells (1997) mention that their model hypothetically could be used with bilingual children there are no studies published in this regard so far. An application would require the availability of normative data on the speech acquisition of the targeted population as well as the availability of assessment tasks in the respective languages. Further, it would require an investigation whether the speech processing process actually proceeds similar in monolingual and bilingual children to ensure the model's validity with bilingual populations.

All of the presented classification systems have their strengths and weaknesses so that internationally no generally accepted classification method exists (Waring & Knight, 2013). What they have in common, though, is that reference to typically developing peers is necessary to categorise the respective child's performance (Dodd, 1995; Stackhouse & Wells, 1997). This is especially important for identifying a delay as this is defined by demonstrating speech behaviours of typically developing younger children (Dodd, 1995; Ingram, 1989b; Shriberg & Austin, 1998).

2.2.1.1 *Speech sound disorders in monolingual German-speaking children*

Recent studies investigating SSD of unknown origin in German speaking children have mainly identified and classified SSD according to Dodd's classification system (Fox-Boyer, 2014a; Fox & Dodd, 2001), apart from a study by Ullrich (2010) which concentrated on diagnosing children within the theory of non-linear phonology.

The most recent data on monolingual German-speaking children with SSD have been reported by Fox-Boyer (2014a). She presented data from the initial assessments of 276 children with suspected SSD or SLI who were referred to an SLT clinic between the years of 2000 and 2011. Children in this retrospective study were aged between 2;8 and 6;11 years and identified as having SSD based on the nature of their phonological patterns. These included patterns that were either considered to be typical but inappropriate for the children's age (i.e., delayed) or those that were regarded as unusual (deviant) for typically developing children. In accordance with other international studies, Fox-Boyer (2014a) demonstrated that the majority of phonological patterns produced by the children represented a delay (Broomfield & Dodd, 2004). In this regard, children demonstrated particular difficulties with the post-alveolar and palatal sibilants /ʃ/ and /ç/ as well as with velar stops and consonant clusters. The most frequent deviant patterns children produced were contact assimilation (/tʁ, dʁ/ → [kʁ, gʁ]), stopping of fricatives and affricates as well as having difficulties with the correct realisation of fricatives in general. The following less frequent deviant patterns were also found (Fox-Boyer, 2014a):

- /ʁ/ → [l]
- /l/ → [j]
- Backing of /t, d, n/
- /f, v/ → [s, z]
- Deletion of all fricatives
- Allophonic use of fricatives
- Deletion of all initial consonant clusters
- Consonant cluster distortion
- Initial consonant deletion
- Deletion of word-final consonant clusters
- Onset process.

In an earlier, smaller-scale study including 100 speech disordered children of which 80 were reported to demonstrate phonological difficulties, Fox and Dodd (2001) reported the following additional phonological patterns to be present: metathesis, intrusive consonants and favourite sound. Ziller and Wohlleben (2006) supplemented this finding by having observed the phonological patterns of fricativation, affrication, nasalisation and denasalisation in 22 German-speaking children aged 4;0 - 4;11 years receiving SLT intervention. This demonstrates that children with SSD form a very heterogeneous group exhibiting at least as many deviant patterns as there are children with SSD (Fox, 2004; Ingram, 1989b). Furthermore, deviant patterns were predominantly found in younger children whereas older children primarily demonstrated delays (Fox-Boyer, 2014a). This may probably be due to the fact that deviant patterns are more salient and create more intelligibility problems which could motivate parents, paediatricians etc. to seek advice from SLTs earlier than for children with a delay (Fox-Boyer, 2014a).

Besides delayed or atypical phonological patterns, an inconsistency rate above 40%, as proposed by Dodd (1995), as well as a reduced intelligibility are considered to be markers for an atypical phonological acquisition in German-speaking children (Fox, 2011; Neumann, Rietz, & Stenneken, 2016).

2.2.1.2 Speech sound disorders in monolingual Turkish-speaking children

For the identification of speech sound disorders in Turkish monolingual children Dodd's (1995) classification system as well as Shriberg and Kwiatkowsky's (1982b) severity rating of PCC-scores have been applied in previous work (e.g., Topbaş, 1997, 2006; Topbaş & Ünal, 2010). Topbaş (2006) identified three types of phonological patterns which resemble those defined by Dodd (1995), i.e., delayed developmental, unusual-consistent and inconsistent-variable patterns. As for German, delayed developmental patterns were produced the most and the frequency of occurrence of these patterns was significantly higher than in typically developing children. Difficulties particularly occurred with the realisation of liquids. Unusual-consistent patterns found in Turkish-speaking children included the following (Topbaş, 2006):

- Stopping of liquids and nasals
- Nasalisation of fricatives
- Gliding of fricatives
- Unusual metathesis
- Glottal stop insertion
- Inconsistent velarisation
- Sound preference
- Lateralisation
- Inconsistent use of /h/.

Topbaş (2006) identified inconsistent productions as patterns that were not found in typically developing children or that occurred variably and idiosyncratically and were somehow definable but difficult to explain systematically by rules. This is in contrast to Dodd's (1995) system in which the identification of an inconsistent speech behaviour requires the administration of a separate assessment in which children are asked to repeat a word three times within one session. From this, their rate of inconsistent productions is calculated and compared to the cut-off for consistent speech (i.e., < 40 % inconsistent realisations; Dodd, 1995).

2.2.2 Markers for SSD in bilingual children

In the limited previous work, SSD in bilinguals have been identified and classified in various ways applying theoretical constructs for monolingual children. Studies by Dodd and her colleagues, for example, used her model for differential diagnosis (Dodd, 1995; Dodd et al., 1997; Holm & Dodd, 1999a, 2001; Holm et al., 1999), whereas Burrows and Goldstein (2010) used PCC-cut-off values for SSD defined by Gruber (1999). The main rationale for taking theoretical constructs for monolinguals as a basis was the insight from cross-linguistic studies which proved their validity for other languages (e.g., Fox & Dodd, 2001; Shriberg et al., 1997a; So & Dodd, 1995; Zhu & Dodd, 2000b). Further, the measures used (e.g., phonological patterns and PCC-scores) had been applied with bilingual children before so that their applicability with these populations was considered to be justified (e.g., Goldstein et al., 2005; Goldstein & Washington, 2001; Holm, 1998). Some projects in which bilingual children have been identified with SSD and the specific phonological markers that were applied are summarised below.

Dodd et al. (1997) examined two Cantonese-English bilingual children who were referred to local SLT clinics on suspect of having speech difficulties. Compared to age-matched typically developing bilingual peers, their participants produced more errors, exhibited other atypical phonological patterns or a general higher number of unusual patterns and distorted phones both in Cantonese and English. Whereas in typically developing peers, phonological patterns were rarely shared across the languages, the opposite was the case for one child in this study. Thus, the markers for identification and differential diagnosis used by Dodd et al. (1997) were of descriptive-linguistic or rather arbitrary quantitative manner. The children's performance was clearly weaker than that of age-matched peers, the cut-off values the authors applied, however, were not defined. Further the number of children in the comparison group (which were part of a separate, preceding study) was very small including only 16 children.
In order to identify SSD in two four-year old Italian-English bilingual children, Holm and Dodd (1999a), compared the children's speech with data of typically developing monolingual children. One of their participants was highly unintelligible due to his inconsistency in realising most phonemes, applying phonological patterns, and producing the same target words during one session. Further his realisations were better

during imitation compared to spontaneous speech (Holm & Dodd, 1999a). This speech behaviour was present in both of his languages but was absent from the speech of his monolingual peers. Holm et al. (1999) found a similar profile in a Pakistani heritage - English bilingual child which was different from his bilingual peers. Given these facts, classifying both children as developing atypically seems reasonable. However, Holm and Dodd (1999a) also used monolingual data as comparison for the identification of SSD in the second child in their study. This child only produced patterns typical for younger monolingual peers and was missing a few later developing phonemes from her inventories in both languages. Classifying her speech performance as delayed might be questionable as no information about the typical Italian-English bilingual speech acquisition could be referred to. Her delay may also be credited to bilingual development which was found to have a decelerating effect on some children (e.g., Bunta et al., 2009; Gildersleeve-Neumann et al., 2008; Salgert et al., 2012). Further, two small-scale studies on Turkish-German bilingual children with (suspected) SSD used the nature of phonological patterns as a marker for an atypical development but applied monolingual norms for the identification of the nature of the patterns (Fox-Boyer et al., 2014; Tugay & Schultz-Ünsal, 2013). The comparison of bilingual speech with normative data of monolinguals, however, may only demonstrate a slower speech acquisition but no clinical delay and further bias/mislead the identification of atypical patterns (Thordardottir et al., 2006).

Burrows and Goldstein (2010), in contrast, used a combination of PCC-scores less than 85% and parental report on the child’s intelligibility to them and to others for the identification of SSD. The bilingual experience including input, proficiency in each language, dialects and frequency of output, however, all varied in their eight Spanish-English bilingual children. Thus, it cannot be clearly stated what exactly influenced the low bilinguals’ PCC-scores. Further, the cut-off for PCC-scores they referred to (i.e., < 85%) was originally identified for monolingual English-speaking children (Gruber, 1999) and had not been examined regarding its reliability and validity for bilingual children.

To sum up, the criteria used to identify speech sound disorders in bilingual children vary across studies and were often of arbitrary manner. Further, they were identified based on the single performance of single cases or small groups so that generalisability is limited (Bedore & Peña, 2008).

Despite the uncertainty regarding how to identify bilingual children with SSD, some studies (mainly those focusing on intervention outcomes) did not state how their children were diagnosed at all. Yavaş (2010) noted that his Spanish-English bilingual participants were diagnosed by SLTs as having a phonological disorder but the type of markers the SLTs used for identification remained unclear. Others only provide information on the speech skills tested shortly before intervention (Gildersleeve-Neumann & Goldstein, 2014; Ramos & Maed, 2014; Ray, 2002) or that children were receiving intervention (Tugay & Schultz-Ünsal, 2013).

In all of the cited studies, children demonstrated weak phonological skills in both of their languages. Dodd et al. (1997) concluded from this that the speech processing deficits underlying the children's SSD have to be universal rather than specific to a language. Further, Holm and Dodd (1999c) discovered that the nature of phonological patterns (i.e., typical, delayed or atypical) was the same across their participant's two languages. This led them to hypothesise that the speech output of bilinguals is generated, restricted and managed by one single underlying mechanism in both of their languages since different natures of phonological patterns imply different underlying speech processing deficits. This finding was confirmed by further small-scale and single-case studies on different bilingual populations (Munro, 1985 cited in Ball, Müller, & Munro, 2006; Dodd et al., 1997; Holm & Dodd, 1999a, 2001; Holm et al., 1999; Lee, Ballard, & Purdy, 2015a) so that researchers generally support the claim that evidence for a delayed or atypical development needs to be present in both of a child's languages for it to be called SSD (Hambly et al., 2013). It has to be noted, though, that one small-scale study on five Russian-German and five Turkish-German bilingual children by Fox-Boyer et al. (2014) contradicts this claim as it revealed that children's performances could not always be allocated to the same subtype of SSD – at least not if basing classifications on monolingual data only. Further, three of their children showed speech difficulties in one language but typical development in the other – a profile generally not attributed with SSD. Weak skills in one language are generally believed to be related to language experience factors (Chilla et al., 2010; MacLeod et al., 2011; Scharff Rethfeld, 2013; Wyatt, 2012). However, in addition to Fox-Boyer et al.'s (2014) findings there are at least two issues with this assumption:

1) If a child's speech is only assessed on one speech processing route (e.g., single-word naming), which was the case in many previous studies, their speech difficulties may not necessarily become apparent in both languages. If a phonological difficulty was, for instance, caused by a subtle weakness in assembling new motor programs this difficulty may be more obvious in the child's weaker/non-dominant language due to a probably lower familiarity with the test items or rather small vocabulary. In their stronger/dominant language, however, this weakness could potentially be masked by their good vocabulary knowledge (Stackhouse & Wells, 1997). Especially, if the child is older and has memorised the correct pronunciation for high frequent words (Pascoe, Stackhouse, & Wells, 2006; Stackhouse et al., 2007). In that case, examinations with further

assessment tasks targeting different speech processing routes and levels are necessary.

2) Children's speech difficulties do not only appear in form of delayed/deviated phonological patterns but in a variety of frequent and infrequent phonological realisations that deviate from the adult target (Dodd, 1995; Stackhouse & Wells, 1997). Thus, if only those variations that occurred frequently enough to be labelled as phonological patterns are considered the degree of variability in a child's speech may be undetected. Consider a child with a large number of infrequent variations from the adult speech model in both languages that additionally shows delayed phonological patterns in only one language. Categorising this child as typical on the basis of delayed phonological patterns only being evident in one language may underdiagnose a delay in achieving consistency of production in both languages – as evident by a large number of variated productions in both languages.

Therefore, it seems even more important to look at different processing levels in bilinguals than in monolinguals. This has also been recommended for the diagnosis of SLI. According to Kohnert (2013) it is very likely that not a single measure will be able to distinguish typically developing bilinguals from atypically developing ones but rather a combination of repeated measures and measures of different levels of speech and language processing. The diagnostic relevance of both are discussed in the following two subsections.

2.2.3 Diagnostic relevance of longitudinal speech data

As children grow older their cognitive abilities maturate and will generally have an improving influence on their speech performances (Weiss, 2007). The relative progress children with SSD show over time may vary significantly from that of typically developing children and will depend on the nature of the difficulty (i.e., delayed versus deviant; Stackhouse et al., 2007; Zhu & Dodd, 2000a). Since children with phonological delays form the lower end of the range for typical development it is rather quantitative than qualitative differences that distinguish them from typically developing children. Children with a phonological disorder, on the contrary, also differ qualitatively (Zhu & Dodd, 2000a). Fox and Brodbeck (2004), for example, investigated the change in phonetic and phonological pattern occurrence of 49 German-speaking children with a phonological delay ($n = 27$) and a consistent phonological disorder ($n = 22$) who were on a waiting list to receive SLT support. Two-thirds of the children with a phonological delay showed a spontaneous recovery from their difficulties after an average waiting period of 8.4

months, whereas two-thirds of children with a consistent phonological disorder continued to show weaknesses after an average waiting time of 5.5 months.

Similarly, Williams and Elbert (2003) conducted a study on the phonological development of five late talkers (with phonological delay) aged between 22 and 31 months at initial assessment and found that those children who used unusual phonological patterns persisted with phonological difficulties at the ages of 33 - 42 months by demonstrating a large sound variability and showing only little progress over time. In contrast, those children who exhibited typical but delayed phonological patterns, were more consistent in the way they used their sounds, showed a larger progress over time and overcame their developmental delays (Williams & Elbert, 2003). Further evidence for a possible spontaneous remission of a phonological delay comes from a study by Roulstone, Peters, Glogowska, and Enderby (2003). They examined 12 English-speaking children with a phonological delay aged younger than 3;6 years. More than 50% of the included children significantly improved their phonological skills to that extent that they no longer met the criteria for a phonological delay after 10 - 12 months. The other children, however, continued to exhibit weak phonological abilities demonstrating significantly more phonological errors (Roulstone et al., 2003). Thus, longitudinally monitoring the qualitative and quantitative changes in children's speech may provide useful diagnostic information. In particular, the clinical decision-making process when diagnosing bilingual children may benefit from such an approach as the normative comparison basis for this population is very small, so clear and generalisable markers for SSD could not be determined yet from a single assessment time-point (Kohnert, 2013; McLeod et al., 2017). It would seem likely that those children whose phonological skills are only weak due to a lack of contact to the respective language improve their skills the more and longer their contact to that language is sustained. Performance might stagnate in a language, though, if its frequency of use decreases over time (Montrul, 2008). Therefore, if collecting longitudinal data it should be ensured that children's language input and output patterns between the measurement intervals are assessed as well and considered during evaluation patterns (Hambly et al., 2013; McLeod, 2012a).

So far, most projects collecting longitudinal data from bilingual children have been conducted with typically developing children (e.g., Anderson, 2004; Gildersleeve-Neumann et al., 2008; Holm & Dodd, 1999c; Johnson & Lancaster, 1998; Keshavarz & Ingram, 2002; Kim et al., 2017; MacLeod et al., 2011; Morrow et al., 2014; Schnitzer & Krasinski, 1994, 1996). Therefore, none of the potential markers identified in bilingual children with SSD has been verified over time. The fact that longitudinal data of typically developing children revealed that children's performances fluctuate over time (i.e., children demonstrate phonological abilities at one point in time which could not be observed in their speech at a later point in time; Anderson, 2004; Kim et al., 2017; Morrow et al., 2014), further complicates the identification of SSD in bilinguals. The "atypical" performances reported for children in studies by Holm et al. (1999), Dodd et al. (1997), and Burrows and Goldstein (2010), for example, could therefore also be due to the typical variability bilinguals show over time (Kim et al., 2017). To account for this, the present research project is of a longitudinal design.

Longitudinal data additionally have the benefit of providing the possibility to evaluate classification models since the change children in the different subgroups undergo over time can be observed (Dodd, 2014). Those data, however, are limited to provide insight to the developmental changes and may not give insight into the nature of the child's speech difficulties. To establish these further, assessments on different processing levels are necessary.

2.2.4 Diagnostic relevance of psycholinguistic speech tasks

Although well-designed single-word naming assessments are a valid and reliable tool to assess monolingual children's phonological output abilities (Bernhardt & Holdgrafer, 2001; Masterson et al., 2005; McLeod & Baker, 2014), atypical performance on this activity may be caused by a variety of factors. A child could have difficulties with top-down processing including, for example, inaccurately stored motor programs or could struggle with bottom-up processing due to hearing difficulties (Stackhouse & Wells, 1997). Therefore, children with the same diagnostic label or similar phonological output may have different psycholinguistic profiles (which contradicts the notion underlying Dodd's (1995) model) so that results of a single test task are not necessarily sufficient to reveal the child's individual level(s) of breakdown in the speech processing chain (Stackhouse & Wells, 1997; Wren et al., 2012). In order to fully fathom a child's difficulties, to distinguish typical from atypical performance or to describe their individual phonological profiles, different test tasks at multiple levels of the phonological hierarchy are recommendable (MacLeod et al., 2011). These seem especially important for diagnosing bilingual children as their assessment results may be confounded by a variety of socio-linguistic factors and no single marker for an atypical development could be found yet (Kohnert, 2013; McLeod et al., 2017). One example is that bilingual children's performance in a single-word naming task might be biased by their lexical (un)familiarity with the words used in the assessment. To avoid this misleading diagnosis the application of additional psycholinguistic tasks which do not require the child to access their lexical representations (e.g., those that include non-words) might be useful (Stackhouse & Wells, 1997) and was pursued in the present research.

In the following subsections three different speech output tasks which do not require the child to access their lexical representations (i.e., the imitation of isolated phones, NWR and consistency in repeating non-words) and have been used in the clinical decision-making process before are described and their (potential) discriminative power in monolinguals and bilinguals discussed.

2.2.4.1 Phone imitation

Imitating phones in isolation is a fundamental task to differentiate whether children's speech difficulties result from a motor execution issue (articulation difficulties) or a different speech processing difficulty (Stackhouse & Wells, 1997). Further, phone imitation is part of a stimulability examination during which children are asked to imitate phones in isolation and in different syllable positions which are absent from their phonetic inventory in spontaneous speech to determine the probability with which these sounds will be acquired by maturation (Powell & Miccio, 1996). Some studies that investigated the stimulability of typically and atypically developing monolingual children highlighted that correct phone imitation is related to typical speech acquisition and evidences that the necessary motoric movement abilities for correct articulation are present (e.g., Powell & Miccio, 1996). In atypically developing children, those sounds that could be imitated in isolation were likely to develop through maturation whereas sounds that could not be accurately imitated were not and needed to be addressed in intervention (Miccio, Elbert, & Forrest, 1999). A child's stimulability is thus assumed to provide useful insights into their speech production abilities and may help to predict the occurrence of speech sounds in a child's phonetic-phonological system (De Castro & Wertzner, 2012). Kubaschk, Fox-Boyer and Klann (2015), however, reported that German-speaking children's inability to imitate phones correctly was not necessarily associated with atypical or delayed phonological development as both a subset of children who only produced age-appropriate phonological patterns were not stimulable for all phones and some children with atypical/delayed phonological patterns were able to imitate all German phones.

Since the variability in bilingual children's speech productions is usually very high (Gildersleeve-Neumann & Wright, 2010; Goldstein & Washington, 2001; Hack et al., 2012; Holm & Dodd, 1999c), it is important to determine whether this is caused by (subtle) motor difficulties (i.e., phonetic disability restricting phonological outcome) or by phonological issues (Hewlett, 1985); especially, if the children's phone imitation skills may predict the unguided learning of speech sounds that are not present in their speech yet.

Although several studies have reported on bilingual children's phonetic inventories based on their performances in spontaneous speech or single-word naming tasks (Fabiano-Smith & Barlow, 2010), there is a lack of studies, in which bilingual children's ability to imitate the phones of their languages in isolation was investigated. Given the relative language-independent nature (Stackhouse & Wells, 1997) and discriminative power of this task in at least some atypical monolingual populations (Miccio et al., 1999) it appears to be a valuable supplement to the process of identifying SSD in bilingual children.

2.2.4.2 Non-word repetition

A child's inability to produce phonologically accurate speech may further result from their difficulty to assemble new motor programs, which in turn is a crucial skill for pronouncing unfamiliar words and extending their expressive vocabulary (Stackhouse & Wells, 1997). A study by Vance, Stackhouse and Wells (2005) has demonstrated that there are significant correlations between typically developing monolingual children's NWR-, real-word repetition and single-word naming skills. Thus, it can be assumed that these tasks reflect children's level of accuracy of speech production in a similar way. Across several studies NWR-skills could be proven to discriminate not only between children with and without SSD (Munson, Edwards, & Beckman, 2005; Shriberg et al., 2009) but also to differentiate between different types of phonological difficulties (Shriberg et al., 2005; Williams & Chiat, 1993), especially when children's performances were compared across speech tasks (e.g., across NWR, real-word repetition, single-word naming). Further, they helped to differentiate persistent SSD from common clinical distortions in eight-year-olds (Wren et al., 2012). Moreover, Stackhouse, Vance, Pascoe and Wells (2007) emphasise that NWR-tasks may have the potential to uncover hidden persisting speech difficulties – those that are not apparent when only real words are produced – as 11 of their 34 psycholinguistically assessed children with persisting SSD performed poorly (i.e., scored lower than -1 *SD*) on NWR-tasks only.

Studies on bilingual children reported that performances on NWR-tasks do not necessarily differ from monolingual peers' performances (e.g., dos Santos & Ferré, 2016; Lee, Kim, & Yim, 2013) but nevertheless often demonstrated influences of language-specific knowledge (e.g., Duncan & Paradis, 2016; Sharp & Mueller Gathercole, 2013; Summers et al., 2010). So, children were found to perform better if the non-word stimuli had phonological features in common with real-words from their respective languages (Chiat, 2015; Sharp & Mueller Gathercole, 2013). Other studies, however, only reported a marginal impact of language experience on bilingual children's NWR-performance (Brandeker & Thordardottir, 2015; Thordardottir & Brandeker, 2013). To rule out effects of language experience, NWR-tasks that are administered in both of a child's languages or language-independent stimuli may be possible solutions. Due to the great variability across the phonological systems of the languages of the world, fully language-independent non-words are not possible to create (Chiat, 2015). However, Chiat (2015) further argues that if a number of factors are considered quasi-language-independent items – i.e., those that are constructed from phonological features that are shared across as many different languages as possible – are designable. Those have been designed and employed in studies within a large European project – Cost Action IS0804, Language Impairment in a Multilingual Society: Linguistic Patterns and the Road to Assessment – on disentangling bilingualism from SLI and have been proven to be discriminative and less susceptible to language experience compared to language-specific NWR-items in bilinguals (Boerma et al., 2015).

The vast majority of studies on NWR-skills in monolinguals and bilinguals was aimed to identify SLI (e.g., Chiat, 2015; Vance et al., 2005). Up to date, no study has been published to investigate the discriminative power of NWR accuracy for the identification of SSD in bilingual children. However, given their usefulness in identifying SSD in monolinguals and the possibility to design them quasi-language-independently, it would be interesting to see if quasi-language-independent non-word stimuli have the potential to predict children's phonological skills in both of their languages and can help to identify children with SSD.

2.2.4.3 Consistency of non-word production

In order to produce consistent (i.e., invariable) speech children's motor planning abilities need to be intact (Williams & Stackhouse, 1998). If these are impaired their speech will be less predictable and therefore difficult to understand (Fox, 2004; Stackhouse et al., 2007). Inconsistent speech productions are very common in the initial stages of children's speech development (Ingram, 1989a; McLeod & Hewett, 2008; Vihman, 2014) but were found to decrease with age (Williams & Stackhouse, 1998, 2000). Therefore, determining if a child's speech is considered to be clinically inconsistent is especially challenging at younger ages and requires normative data for comparison (Dodd, 1995; Stackhouse & Wells, 1997).

Williams and Stackhouse (1998), having assessed monolingual English-speaking children, reported that children with SSD demonstrated different error quantities and qualities in the consistency tasks compared to typically developing children. Further, even atypically developing children demonstrated different performances on the consistency measure although their overall number of errors in the single-word naming task was comparable. This suggests that consistency skills may have sufficient power to discriminate between typically and atypically developing monolingual children (Williams & Stackhouse, 1998). Although measuring consistency skills slightly differently, Dodd (1995) also demonstrated their potential to discriminate between typical and atypical development as well as between different forms of speech sound disorders.

Since several studies have reported that bilingual children produce a large number of phonological patterns (Goldstein & Washington, 2001; Hack et al., 2012) the question is what drives this intra-individual variability. Do bilingual children generally have weak motor planning skills or only if developing atypically? If so, the question is whether inconsistency scores could help to predict children's accuracy in single-word naming tasks.

Grech and Dodd (2008) are one of the few reporting bilingual children's typical performance on a consistency task. This task was designed on the basis of the 25-word inconsistency test included in the *Diagnostic Evaluation of Articulation and Phonology* (DEAP; Dodd, Zhu, Crosbie, Holm, & Ozanne, 2002). Grech and Dodd's (2008) Maltese-English bilingual participants aged between 2;0 and 6;0 years demonstrated age-related performances with younger children being more inconsistent than older children, showed an effect for the language input at home (i.e., either monolingual Maltese or bilingual Maltese-English) and generally achieved scores below the 40% cut-off score for inconsistency.

Furthermore, inconsistency tasks have also rarely been reported to be used for differential diagnosis in bilingual children. Holm and Dodd (1999b), for example, used an English inconsistency assessment (the 25-words inconsistency test of the DEAP; Dodd et al. 2002) as well as a set of 20 words from the Rochdale Assessment of Mirpuri Phonology (Stow & Pert, 1998) which had to be named three times within one session to examine the consistency rate of a Punjabi-English bilingual child with suspected SSD. However, these assessments had not been validated for bilinguals so that their discriminative accuracy remains unknown. Thus, in order to explain the large variability in bilingual children's phonological performances and to investigate this potential marker for SSD in bilinguals, consistency skills need to be investigated in a larger sample and with appropriate stimuli (e.g., language-independent non-words to reduce a bias by vocabulary knowledge). Both aspects were intended to be addressed with the design of and the materials included in the present research.

2.2.5 Summary – Identification of SSD in children

The literature reviewed in this section (2.2) has highlighted that data on typical and atypical development is necessary in order to reliably identify SSD in children and to determine clinical markers (Dodd, 1995; Stackhouse & Wells, 1997). For monolingual children, three major classification systems have been developed but there is no international consensus on which of them (and their associated clinical markers) has the best evidence (Waring & Knight, 2013). Even though there are several studies on bilingual children with (suspected) SSD, these only include single-case and small-scale projects and thus do not provide the normative basis required for a clear specification of clinical markers in this population. As a consequence, classification systems and clinical markers established for monolinguals have been used with bilinguals but their validity remains unclear and requires further investigation with larger samples.

As a supplement to the traditional single instance confrontation naming examination, longitudinal data as well as results from different psycholinguistic tasks have been proven to be useful in the decision-making process in monolinguals (Roulstone et al., 2003; Stackhouse & Wells, 1997). Their relative importance for the phonological acquisition in bilingual children, however, is unclear and warrants further exploration which is addressed in the present research project.

Overall, Chapter 1 and 2 have highlighted that the typical and atypical phonological acquisition in bilinguals is essentially characterised by the two linguistic systems in contact and the child's individual experiences with them. Thus, generalisations to other language combinations are limited to a certain extent and require the direct investigation of the respective language pair. Due to the lack of larger-scale studies on Turkish-German bilingual children's typical and atypical phonological acquisition its comprehensive investigation is intended with this thesis. The specific aims, research questions and hypotheses postulated for this project are outlined in the following.

2.3 Research aims, questions and hypotheses

Due to the world-wide increase in migration (United Nations, 2016) the number of bilingual speakers (adults and children) is growing constantly. It is well known that children growing up bilingually perform differently to their monolingual peers on language (i.e., phonological) development (see Section 1.2) and that the differentiation of typical and atypical development is not a simple task (see Section 2.1.2). This is, however, of great importance since the availability of equal educational chances needs to be guaranteed for bilingual children as well and because speech and language problems can interfere with these chances (McCormack et al., 2009; Nathan et al., 2004b). The different developmental paths in bilingual children in comparison to monolingual children are caused by the challenging task of acquiring two or more languages at the same time or soon after each other as well as the heterogeneity in the levels of experience and proficiency in each language (Goldstein et al., 2010; Paradis & Genesee, 1996). Additionally, it is well known that the languages to be acquired influence the course of development (e.g., cross-linguistic transfer; MacWhinney, 2005b; Paradis & Genesee, 1996) so that it is not possible to investigate and compare children acquiring different languages within one group.

In order to understand the phonological development in bilinguals of a specific language combination and to differentiate typical from atypical development in this population normative data are required. Furthermore, test material to investigate typical development needs to be available. However, both is rarely the case (McLeod et al., 2017). Despite Turkish-German bilingual children forming the largest group of non-native speakers in Germany, characteristics of their typical and atypical speech acquisition are not fully researched and understood to date. Further, no reliable or valid assessment tools are available, and thus the identification of at-risk children remains particularly challenging (McLeod et al., 2013; Williams & McLeod, 2012; Yavaş, 1998). Previous studies on monolinguals have shown that observing children's speech development beyond a single assessment point does not only provide insight into its qualitative and quantitative changes over time but may also shed light on markers for an atypical development (Roulstone et al., 2003; Williams & Elbert, 2003). This aspect seems especially important for the evaluation of bilingual speech development due to the children's usually constantly changing input and output situations and high variability in speech productions (Kim et al., 2017). However, longitudinal data in three- to five-year-olds are very scarce and have not been collected for Turkish-German bilingual children living in Germany yet.

In addition, studies on other aspects of language acquisition in bilingual children have revealed that it will probably always be a combination of several assessments that provides a clear diagnosis of children's speech and language abilities (Kohnert, 2013). Thus, the isolated investigation of children's single-word naming skills may be insufficient. This view has also been held by representatives of the psycholinguistic approach who recommend the inclusion of a variety of psycholinguistic assessment tasks to create an individual profile of a child's strengths and weaknesses in the speech processing chain that might facilitate the differentiation of typically from atypically developing children (Stackhouse & Wells, 1997). Bilingual children's speech, however, has only marginally been investigated in this respect (cf. Grech & Dodd, 2008).

Consequently, the application and reliability of psycholinguistic tasks in bilinguals has been insufficiently researched so far.
Hence the following two main research aims for the present research have been set:

1) **To describe the typical phonological acquisition in Turkish-German bilingual children by means of valid and reliable test instruments**
2) **To identify markers that make it possible to differentiate typical phonological development from atypical development in Turkish-German bilingual children**

To achieve aim 1 the following research questions have been posed:

1) How does the phonological acquisition in Turkish-German bilingual children proceed in both of their languages?
   a) At what rate do Turkish-German bilingual children acquire the consonants and consonant clusters of both of their languages?
   b) What type of phonological variations do Turkish-German bilingual children show in either/both of their languages?
2) Do age, gender, input and output patterns as well as language proficiency have an influence on children's phonological performances? And if so, in which way?
3) How does the phonological acquisition in Turkish-German bilingual children progress over a course of 12 - 15 months?
4) To what extend does the knowledge of the phonological development in Turkish-German bilingual children and its influencing factors contribute to theoretical discussions on phonological development in general and in bilingual children specifically?

Based on the experiences from pilot studies (Salgert et al., 2012; Ünsal & Fox, 2002), cross-linguistic theoretical explorations (Fabiano-Smith & Goldstein, 2010b; Paradis, 2001) as well as the assumption that bilingual children's speech development is influenced by the respective languages in contact as well as further socio-linguistic factors the following hypotheses were formulated:

a) Turkish-German bilingual children's rate of acquisition will mainly differ from that in monolingual children demonstrating evidence for a slower, faster and similarly fast rate of acquisition due to an interaction of a bilingual's two languages (e.g., Paradis & Genesee, 1996).

b) Turkish-German bilingual children will show phonological patterns similar and dissimilar to those reported for monolinguals of either language as well as evidences for CLT because of an interaction of a bilingual's two languages (e.g., Holm & Dodd, 1999c).

c) Turkish-German bilingual children's phonological acquisition should be influenced by:
   - age as maturation has a positive effect on children's general development (Kirk & Vigeland, 2014; Weiss, 2007).
   - gender (at least marginally) since boys' phonological performances have frequently been reported to be weaker than girls' (Dodd et al., 2003; Ege, 2010; Schäfer & Fox, 2006).
   - language input and output patterns as a greater language experience offers more occasions for practising phonological perception and production skills (Davis & Bedore, 2013; Pearson et al., 1997).
   - language proficiency since the individual language domains interact during speech and language development (Davis & Bedore, 2013; Kehoe, 2011).

d) The phonological acquisition in Turkish-German bilingual children is suggested to generally progress in both of their languages but the improvement will differ across languages, since bilinguals rarely have balanced skills in both of their languages (Grosjean, 1989). Further, individual performances may suggest a stagnation or regression in one or both of the children's languages that is triggered by the acquisition of new knowledge and an associated reorganisation of their linguistic systems (Kim et al., 2017).

e) Knowledge of the Turkish-German bilingual phonological acquisition and its influencing factors will:
   - provide insight into the role of ambient language phonology as the different patterns of ease and challenge in a language can be studied within one individual (Davis & Bedore, 2013).
   - enhance our understanding of variability during children's speech acquisition since language and acquisition contexts in bilinguals vary significantly from one to the other (Pearson, 2007).

To achieve aim 2 the following research questions have been posed:

5) Which factors, quantitative or qualitative, can be identified to support the differentiation between typical and atypical phonological development in Turkish-German bilingual children?
   a) Can markers for a typical and atypical development be identified in actual and longitudinal phonological skills? If so, what type?
   b) Can language-independent psycholinguistic skills predict language-specific single-word naming outcomes in Turkish-German bilingual children and therefore support the identification of SSD?

Based on the literature reviewed as well as the assumption that a SSD will be present in both of a bilingual child's languages (Dodd et al., 1997) the following hypotheses were formulated:

a) The nature of phonological patterns is assumed to function as a clinical marker for SSD as evidenced cross-linguistically with Dodd's (1995) classification system.
b) Children's number of infrequent variants are required to support the marker 'nature of phonological patterns' since they could provide insight into children's developmental level of speech production (Fox, 2016; Stackhouse & Wells, 1997) and help to consider qualitative as well as quantitative aspects of children's speech skills.
c) Longitudinal data will further give insight in the stability of categorisation based on the nature of phonological patterns and number of InfVar over time. Typically developing children will continue to show normal speech performances over time whereas atypically developing children either improve (overcome a delay) or stagnate in their acquisition since the progression of phonological skills over time is depending on the type of SSD (Fox & Brodbeck, 2004; Zhu & Dodd, 2000a).
d) Children's performances on all psycholinguistic tasks are anticipated to predict their performances on single-word naming skills in both languages at t2 since the applied psycholinguistic tasks require to some extent the same speech output processing skills (Stackhouse & Wells, 1997; Vance et al., 2005).

3 METHOD AND PROCEDURES

This chapter presents the applied methods and procedures to achieve the aims outlined in Section 2.3. Section 3.3 on tasks and materials also includes a detailed description of the design and the rationale for the Turkish phonology assessment for Turkish-German bilingual children designed for this research project.

3.1 Study design

In a cohort study of a community sample the children's phonological skills in German and Turkish were assessed once (t1) and followed-up 12 to 15 months after the initial test session (t2; see Figure 3.1).

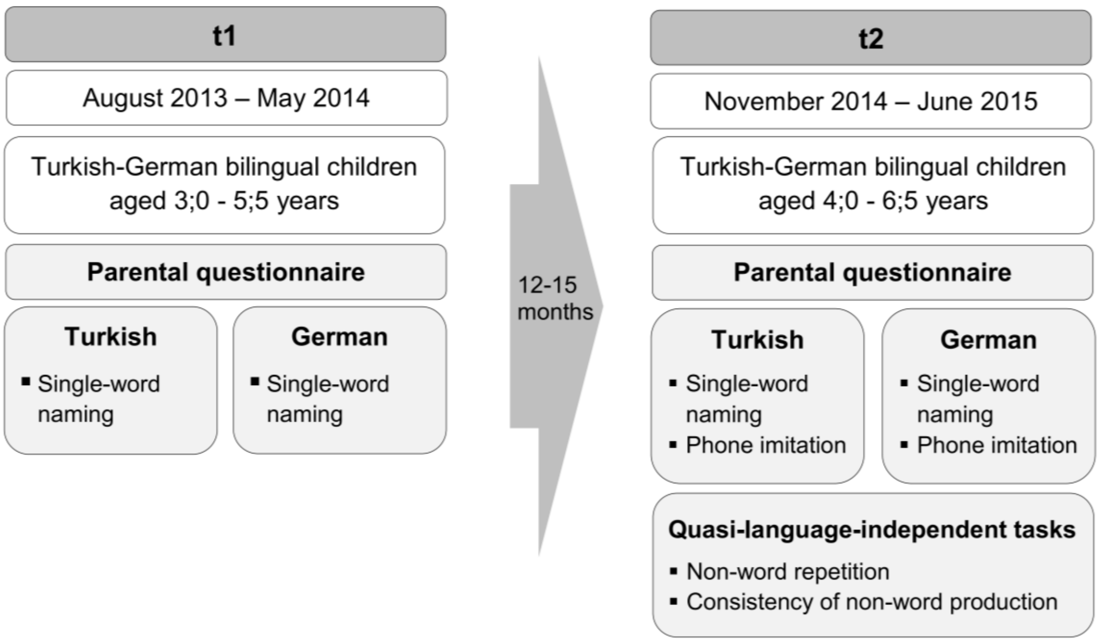


Figure 3.1: Study design

Assessments at t1 were used to collect data on Turkish-German bilingual children's typical phonological acquisition. These are a prerequisite for discriminating typically from atypically developing children (McLeod et al., 2013; Prezas et al., 2014). The second time-point resulted from the analysis of t1-data (i.e., was not originally planned and required re-recruitment of nurseries and participants). Data collected at this time-point served to track the children's progress in their speech development over time and to obtain more specific information on the children's speech processing skills by including additional psycholinguistic tasks. The investigation of children's progress over time is particularly important to understand the shape, direction and rate of Turkish-German bilingual children's phonology growths trajectories (Iglesias & Rojas, 2012). In addition, parental questionnaires were applied at both assessment points to gather comprehensive background information on the children's bilingual situation.

The ethics review panel in the Department of Human Communication Sciences within the University of Sheffield has ethically approved this research project in line with the University's ethics review procedures (see Appendix B and Appendix C).

3.2 Participants

3.2.1 Recruitment for t1 and t2

To reflect the typical population of Turkish-German bilingual children living in Germany as far as possible, the recruitment took place via nurseries in different parts of Germany (i.e., Hamburg, Lübeck and Remscheid). These cities varied in population ranging from 111,766 inhabitants in Remscheid (Stadt Remscheid, 2016) to 1.78 million in Hamburg (Statistisches Amt für Hamburg und Schleswig-Holstein, 2016), and nurseries were located in different parts of the cities.

Head teachers of nurseries were contacted via e-mail to explain the research project and its aims. A week later, a follow-up phone call was scheduled to find out whether the head teachers and their nurseries were interested in taking part. This also allowed for an explanation of the research project and a discussion of any queries. Interested head teachers of nurseries were asked to give their consent to the study being conducted in their institution and subsequently to distribute information leaflets, consent forms and questionnaires to parents whose children met the following selection criteria at t1:

- growing up bilingually with Turkish and German (i.e., using both languages in everyday communication)
- aged between 3;0 and 5;5 years
- having had at least seven months of contact to the German language (e.g., being in nursery for seven months) to ensure a sufficient familiarity (e.g., vocabulary) with this language
- developing typically (i.e., no known syndromes or cognitive difficulties)
- no diagnosed hearing disorders or speech and language difficulties
- not having been referred to speech and language therapy.

Two-hundred-seventy-one parents were contacted via 25 nurseries. Ninety-four parents from 22 nurseries returned signed consent forms and filled-in the parental questionnaires. However, ten children had to be excluded as they did not meet the selection criteria or were not able to/refused to name all test items in at least one language. Hence, the final study sample consisted of 84 children from 20 nurseries at t1 (see Figure 3.2).

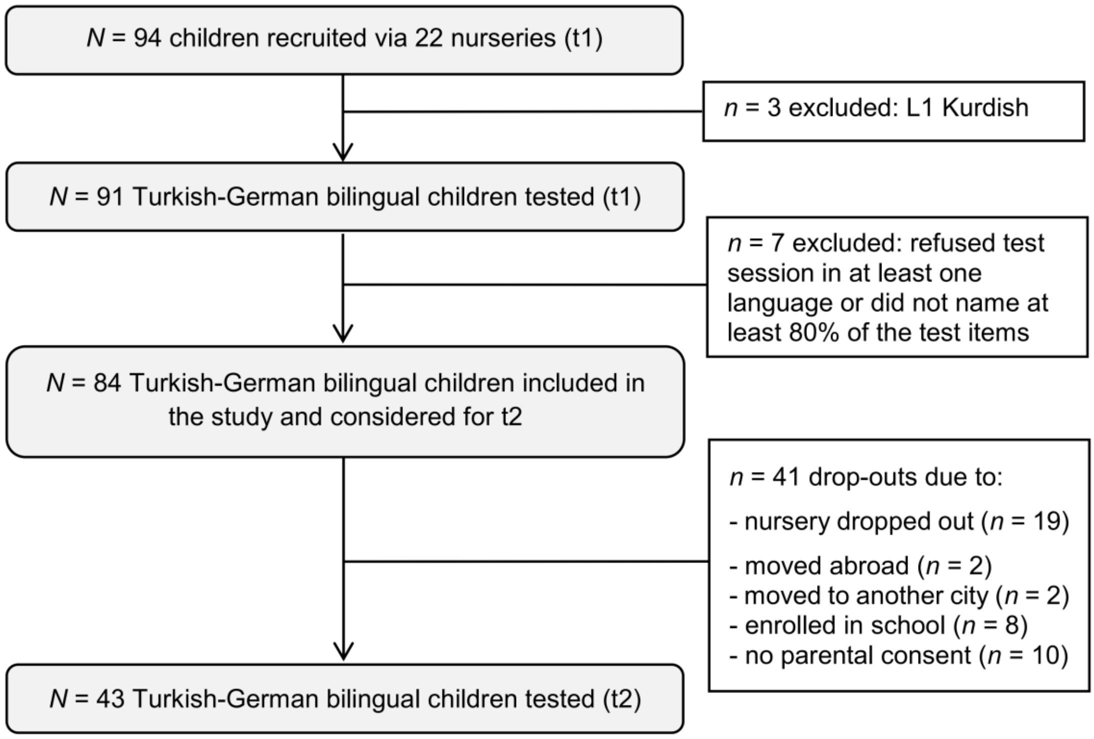


Figure 3.2: Flow of participants through longitudinal project

Fifteen head teachers of t1 nurseries gave their permission for a second test phase (t2) and distributed the information leaflets and consent forms to the respective parents from t1. Consequently, 19 children could not be re-contacted for the follow-up assessment. From the remaining 65 parents 22 could not be re-recruited for a variety of reasons (see Figure 3.2). Thus, the number of participants at t2 comprised 43 Turkish-German bilingual children from 14 nurseries, which is about 51% of the sample at t1.

Assessments at t2 were aimed to take place 12 months after the initial assessment sessions but were dependent on the children's and testers' availability (e.g., due to holidays, illness, nursery activities). On average, t1 and t2 were 13.7 months apart (range: 12 - 15.5 months).

3.2.2 Sample composition at t1 and t2

Detailed information about gender and age distribution at t1 and t2 is displayed in Table 3.1. At both assessment times, children were separated into five 6-months age bands. Female participants outnumbered male participants and the number of children per age group was not evenly distributed at t1 and t2.

Table 3.1: Gender and age distribution of the project's samples at t1 and t2

Age group	*N*		Male		Female		M_{age} (months)		SD_{age} (months)	
	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
3;0 - 3;5	12		6		6		38.0		1.6	
3;6 - 3;11	21		8		13		45.1		1.7	
4;0 - 4;5	17	4	6	0	11	4	50.8	51.5	2.0	1.7
4;6 - 4;11	27	9	12	5	15	4	56.4	57.2	1.8	1.8
5;0 - 5;5	7	11	5	3	2	8	61.1	62.5	0.8	1.8
5;6 - 5;11		13		8		5		68.2		1.6
6;0 - 6;5		6		3		3		74.5		1.4
Total	**84**	**43**	**37**	**19**	**47**	**24**				

As assessments at t2 could not be scheduled exactly twelve months after t1 for all participants, some children had shifted more than two age groups at t2. Figure 3.3 presents the exact age group shift for the 43 participants tested at both assessment times.

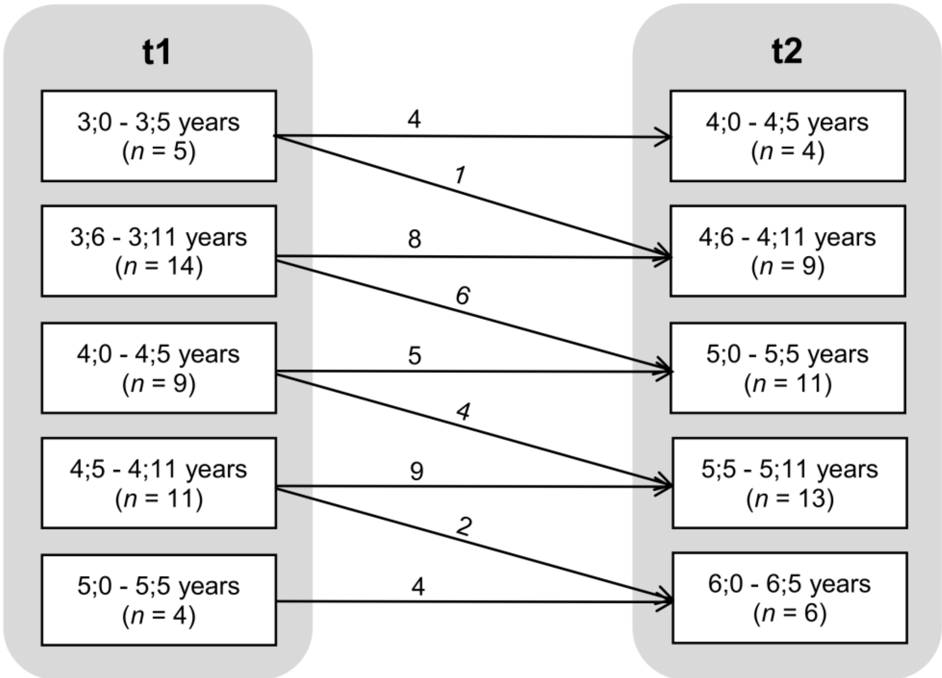


Figure 3.3: Participants' age group shift from t1 to t2 (N = 43)

3.2.3 Parental questionnaire data at t1 and t2

To obtain background information on the children's bilingual development and current bilingual situation parental questionnaires were distributed prior to both assessment time-points (see Appendix D and Appendix E). These were designed based on questions/aspects generally raised in the relevant literature but especially those included/addressed by Gutiérrez-Clellen and Kreiter (2003), Goldstein and Bunta (2012), and McLeod (2012a). Further, some questionnaire data (i.e., gender, language input/output patterns and language proficiency) were used for an analysis of influential factors on children's phonological acquisition (see Section 4.5). However, due to the small sample size this analysis was only carried out for t1 data. At t2, questionnaires were mainly used to identify general tendencies for a change in children's socio-linguistic situations, especially their exposure to Turkish and German (input), their language usage patterns (output) and language proficiency since t1 (see Section 3.2.3.3).

Further, questionnaires at t2 were applied to gather information about whether children had received/were receiving speech and language intervention in the time between t1 and t2. If so, parents were asked to provide details on the start and end date of the therapy, the SLT's contact details as well as to give permission to the researcher to contact the child's SLT. Contacting the SLT was necessary to gather details about the focus of intervention the child had received, since this aspect was particularly valuable for the interpretation of the children's speech abilities at t2.

For the purpose of this thesis only the following background variables will be presented: parents' first languages, age of onset of Turkish and German, language exposure and usage patterns, variation of dialects in the child's regular input, language proficiency and SLT support. Although an analysis of further aspects of the questionnaire data would have been desirable, this was beyond the scope of this thesis.

3.2.3.1 Parents' first language(s)

Most parents indicated that their first language was Turkish, however, some reported that one of them spoke an additional language/dialect or had an L1 other than Turkish (cf. Table 3.2). All parents spoke German as an additional language.

Table 3.2: Parents' first language(s) based on parental questionnaires at t1

L1 / Parent	Turkish *n* (%)	Turkish-German *n* (%)	German *n* (%)	Turkish-Kurdish *n* (%)	Kurdish *n* (%)	Zaza-Turkish *n* (%)
Father	75 (89)	4 (5)	2* (2)	2 (2)	1 (1)	1 (1)
Mother	78 (93)	3 (4)	2* (2)	0 (0)	0 (0)	1 (1)

Note: ***These mothers and fathers did not belong to the same families.

3.2.3.2 Age of acquisition

Sixty-three percent of the parents (*n* = 53) reported that their child acquired Turkish from birth and German later in life, but the latest at age three. For two children parents indicated that children were first exposed to German at age four. An additional child started acquiring Turkish at age two and German from birth. These 56 participants were regarded as sequential bilinguals (Meisel, 2004)[8]. For only 33% of the children (*n* = 28) parents reported that they were exposed to both languages from birth. They were categorised as simultaneous bilinguals (Kohnert, 2013). Since an AoA between the ages of 0 and 3 years was not found to have an effect on children’s phonological performances (Goldstein et al., 2010; Holm & Dodd, 2006) and no differences between simultaneous and sequential bilinguals could be identified on any measure in the present research, data are presented for both groups combined throughout this thesis.

3.2.3.3 Language input and output patterns

At both assessment time-points, all parents indicated that their children used Turkish and German in their everyday lives. The degree of precision in their responses, however, varied considerably. Whereas most of them provided an exact number of hours of input and output per language, others responded in a vague and questionable manner or not at all for either one or both languages. Vague responses included time specifications, such as “half a day” or “often”. Questionable were those responses where parents indicated that their child was exposed to and spoke both languages 24 hours per day. Vague (German input: *n* = 4; German output: *n* = 6; Turkish input: *n* = 3, Turkish output: *n* = 3), questionable (German input: *n* = 0; German output: *n* = 0; Turkish input: *n* = 3, Turkish output: *n* = 2) and no responses (German input: *n* = 6; German output: *n* = 7; Turkish input: *n* = 6; Turkish output: *n* = 8) could not be validly transformed into a single number of hours, thus were excluded from group analyses. Table 3.3 presents the

[8] Originally referred to as *early second language learners* in this publication.

average number of hours of all analysable responses – even though they might have only been provided for one of the child's languages.

Due to parents' apparent difficulty to accurately estimate their children's amount of hours of language input and output, the indicated hours were used to determine the proportion of input/output in German to input/output in Turkish. This way the data could be used for further analyses (influential factors on children's phonological performances).
Three different input/output categories were formed for which only children were considered for whom responses had been provided precisely in both languages: (a) more input/output in German, (b) equal input/output in both languages, and (c) more input/output in Turkish (see Table 3.3). The difference between the indicated hours in German and Turkish needed to be more than two hours for a child to be allocated to either the more input/output in German (moreG) or the more input/output in Turkish (moreT) category. This arbitrary cut-off was used to account for the expected degree of inaccuracy in the parents' responses, since indicating the exact number of hours their child was exposed to a certain language per day was considered to be rather challenging (Goldstein et al., 2010).

Table 3.3: Average input and output patterns in Turkish and German at t1

Language		Input	Output
German	*n* (%)	74 (88)	71 (85)
	M_{Hours}	7.7	7.2
	SD_{Hours}	2.5	2.7
	$\text{Range}_{\text{Hours}}$	4 - 16	1 - 16
Turkish	*n* (%)	72 (86)	69 (82)
	M_{Hours}	6.6	6.1
	SD_{Hours}	3.1	3.6
	$\text{Range}_{\text{Hours}}$	0.5 - 14	0.5 - 16
Proportion	MoreG (*n,* %)	22 (26)	24
	Equal (*n,* %)	33 (39)	29
	MoreT (*n,* %)	17 (20)	16

Note: **moreG**: proportionally more language input/output in German than in Turkish, **equal**: equal amount of language input/output in both languages, **moreT**: proportionally more language input/output in Turkish than in German

As can be seen from Table 3.3, participants received slightly more input and produced slightly more hours of output per day in German ($Mdn_{Input} = 7.8$, $Mdn_{Output} = 7.0$) than in Turkish ($Mdn_{Input} = 6.0$, $Mdn_{Output} = 6.0$). Wilcoxon signed-rank tests, however revealed no statistically significant differences (Input: $z = -1.57$, $p = .059$, $r = -.19$; Output: $z = -1.05$, $p = .149$, $r = -.13$). The child for whom only one hour of output per day in German was reported attended a German nursery for six hours a day. Thus, it could be assumed that the actual number of hours of output was higher than reported by the parents.

As Table 3.4 displays, for most children their amount of language input and output had changed for one or both of their languages by t2. However, on the group level, none of the differences between t1 and t2 were statistically significant according to Wilcoxon signed rank tests:

- German input: $Mdn_{t1} = 8.00$, $Mdn_{t2} = 8.00$, $z = -.52$, $p = .307$, $r = -.09$
- German output: $Mdn_{t1} = 7.00$, $Mdn_{t2} = 7.50$, $z = -.50$, $p = .312$, $r = -.09$
- Turkish input: $Mdn_{t1} = 6.50$, $Mdn_{t2} = 5.00$, $z = -1.01$, $p = .162$, $r = -.20$
- Turkish output: $Mdn_{t1} = 6.00$, $Mdn_{t2} = 5.00$, $z = -1.20$, $p = .118$, $r = -.23$.

Table 3.4: Comparison of children's input and output patterns at t1 and t2 (*N* = 43)

Question	German				Turkish			
	t1 > t2 *n* (%)	**t1 < t2** *n* (%)	**t1 = t2** *n* (%)	**NR**	**t1 > t2** *n* (%)	**t1 < t2** *n* (%)	**t1 = t2** *n* (%)	**NR**
Hrs input per day	16 (37.2)	10 (23.3)	12 (27.9)	4	17 (39.5)	12 (27.9)	5 (11.6)	9
Hrs output per day	9 (20.3)	16 (44.2)	12 (32.6)	5	17 (39.5)	10 (23.3)	6 (14.0)	10

Note: **NR**: no (analysable) response

If the amount of input and output in German and Turkish were compared, children's language exposure and use changed in favour for German as this time, Wilcoxon signed-rank tests revealed highly significant differences between the two languages: Input: $z = -2.95$, $p = .001$, $r = -.49$; Output: $z = -2.87$, $p = .002$, $r = -.47$.

3.2.3.4 Exposure to variations of Turkish

According to the parents' reports many children were exposed to Turkish and German dialects at t1 and t2. Those included Aegean ($n = 1$), Mediterranean Sea, also known as *Akdeniz* ($n = 1$), (Central-) Anatolian ($n = 4$), Black Sea ($n = 1$) and Istanbul dialect ($n = 5$) for Turkish. Twelve parents indicated that their child was exposed to no Turkish dialect as they would speak "Normal/Standard Turkish". For German, the Northern German ($n = 72$), Hanoverian ($n = 1$) and Rhenish dialect ($n = 9$) were reported. However, the majority of parents ($n = 60$) did not respond to this question for Turkish and two did not respond to this question regarding German dialects.

3.2.3.5 Language proficiency

Table 3.5 shows that parents rated their children's language proficiency more frequently to be very good in Turkish than in German at t1. Also, the number of children for whom satisfactory or poor language skills in German were reported was larger than for Turkish.

Table 3.5: Parents' rating of their child's language proficiency in both languages at t1

German \ Turkish	1	1 - 2*	2	2 - 3*	3	3 - 4*	4	Total
1 very good	8	0	6	0	0	0	2	16
1 - 2 very good – good*	0	1	0	0	0	0	0	1
2 good	11	1	13	0	3	0	2	30
2 - 3 good – satisfactory*	2	0	0	1	1	0	0	4
3 satisfactory	16	0	5	0	1	0	1	23
3 - 4 satisfactory – poor*	1	0	0	0	0	1	0	2
4 poor	3	0	3	0	1	0	0	7
Total	41	2	27	1	6	1	5	83

Note: *****Most parents clearly indicated their child's language proficiency in both languages but some ticked two proficiency groups instead to indicate a range.

Overall, 25 parents (29.8%) rated their child's language proficiency to be equal in both languages at t1. Forty-three parents (51.2%) indicated their child was more proficient in Turkish than in German and 15 (17.9%) reported the reverse. One parent (1.2%) did not report their child's language proficiency in German but indicated their language proficiency in Turkish as poor (not listed in Table 3.5).

Similar to the language input and output situation at t2, some children's language proficiency changed from t1 to t2 as well (as rated by their parents, see Table 3.6). In many cases, children's language proficiency in German improved whereas their proficiency in Turkish either remained or deteriorated compared to t1.

Table 3.6: Comparison of parents' rating of their children's language proficiency in German and Turkish at t1 and t2 (*N* = 43)

Language	t1 > t2	t1 < t2	t1 = t2	NR
German	5 (11.6%)	19 (44.2%)	18 (41.9%)	1
Turkish	15 (34.9%)	4 (9.3%)	24 (55.8%)	0

Note: **NR**: no (analysable) response

3.2.3.6 Speech and language intervention between t1 and t2

Two parents indicated that their child was on a waiting list for speech and language therapy but had not received any treatment since t1. Five parents indicated that their child had received or was still receiving speech and language therapy since the initial assessment session. These children were aged 4;10 (*n* = 2), 5;7, 5;9 and 5;11 years at t2. Two parents of these five children allowed the researcher to contact the respective speech and language therapist (SLT). The SLTs were approached to find out about the correct start and end date of the therapy, the number of therapy sessions the child had received between t1 and t2, the language area the SLT's focused on in this time, the language the therapists used and the treatment approach they the applied. A summary of the SLTs' responses can be found in Appendix F.

3.3 Assessment tasks and materials

To investigate the phonological acquisition at t1 and t2 single-word naming phonology tests in both languages were applied. These were supplemented by phone imitation, non-word repetition (NWR) and consistency of non-word production tasks at t2 which were applied to obtain additional psycholinguistic output data that would help to understand the children's phonological profiles and to facilitate a differentiation between typically and atypically developing children. The following subsections describe the individual assessment tasks and diagnostic tools used as well as their reliabilities.

3.3.1 Single-word naming

To assess children's phonological skills single-word naming assessments were applied in both languages. Despite an ongoing debate in the literature about how children understand, control, save and produce speech (Baker, Croot, McLeod, & Paul, 2001) and the known disadvantages of single-word naming assessments as opposed to spontaneous speech data (Klein & Liu-Shea, 2009; Morrison & Shriberg, 1992; Stoel-Gammon & Dunn, 1985), single-word naming has also been found to validly measure children's speech abilities (Bernhardt & Holdgrafer, 2001; Masterson et al., 2005; Wolk & Meisler, 1998) if they were carefully designed (see Section 2.1).

A standardised phonological assessment tool designed for German-speaking children, the Psycholinguistische Analyse kindlicher Aussprachestörungen – PLAKSS II (*Psycholinguistic Analysis of Children's Speech Disorders;* Fox-Boyer, 2014c), was used to assess German phonology skills. Since all participating children were growing up in Germany it was assumed that the quality of German input they received through nursery and the environment (i.e., German friends, hobbies, in public etc.) was likely to resemble that of monolingual German-speaking children. Only the German input children might receive in their home environment was presumed to be of different quantity and quality. Thus, an assessment designed for monolingual children of the environmental language was considered appropriate; especially as currently no single-word naming test exists which is specifically designed to assess German speech skills in Turkish-German bilingual children.

For Turkish, however, it was assumed that the children received a somewhat different input from monolingual Turkish-speaking children in Turkey (e.g., due to the absence of input through nursery and the public environment as well as because of an interaction of Turkish with the environmental language, German). Standardised phonology assessments for Turkish-speaking children living in Turkey were not considered to be appropriate as their included vocabulary was considered to be inappropriate for Turkish-speaking children growing up in Germany (cf. Rinker, Budde-Spengler, & Sachse, 2016). The few published phonological assessment tools for Turkish-German bilinguals (i.e., WIELAU-T (Lammer & Kalmár, 2004), TAT (Naş, 2010) and SCREEMIK 2 (Wagner, 2008)) seemed to be inappropriate for reasons outlined in Section 2.1.3. Therefore, a new set of items which was likely to resemble the vocabulary of young Turkish-German bilingual children growing up in Germany and that met the internationally required criteria for speech assessments was designed for this research project (cf. Section 2.1.1).

3.3.1.1 Design of the Turkish Phonology Assessment for Turkish-German bilingual children

A pilot version of the Turkish Phonology Assessment for Turkish-German bilingual children living in Germany (TPA) was designed in 2011 by Salgert considering the following criteria:

1) Every Turkish consonant should be assessed at least two times in at least two different positions of the word.
2) Every Turkish vowel should be tested at least twice.
3) Mono- to trisyllabic words should be included due to their diagnostic significance (James et al., 2008b) and because Turkish-German bilingual children are likely to be used to polysyllabic words from an early age since Turkish is an agglutinative language and disyllabic words are preferred over monosyllabic words during early acquisition (KopkallI-Yavuz & Topbas, 2000).
4) At least one early, middle and late acquired consonant cluster (cf. Fabiano-Smith & Goldstein, 2010a; Topbaş, 2007) should be assessed.

5) Each typical and atypical phonological pattern for monolingual Turkish-speaking children should have the chance to occur at least three times in the item set.
6) Words with phonotactic constructions that do not occur in German (e.g., voiced fricative in coda position) should be included at least three times each to investigate how children deal with conflicting phonotactic constraints in the two languages.
7) Words that exist in Turkish and German with a similar pronunciation but different stress patterns should be included (e.g., /ˈteːləfon/ and /tɛlɛˈfon/) to identify how bilingual children cope with this similarity and whether they had differentiated phonological systems.

This 55 item set was then discussed with a native speaker regarding the vocabulary and pictures included as well as potential dialectal variations in the phonology of the selected words. Subsequently, the item set was used in a small pilot study with 19 Turkish-German bilingual children aged 2;10 - 3;6 and 3;10 - 4;6 years (Salgert, 2011). The results of the pilot study, however, highlighted some lexical difficulties and picture recognition issues with the chosen items in the children.

Moreover, there were too few occurrences of some phonemes and insufficient sampling of words to explore some types of common and uncommon phonological patterns to the current standard (i.e., at least five opportunities for a pattern to occur; cf. Dodd, 2005). Thus, for the revision of the TPA the following linguistic and cultural criteria were considered to provide a sensitive test to investigate Turkish-German bilingual children's phonological abilities in Turkish:

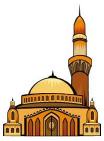
1) All consonants as well as frequent allophones occur at least five times as a singleton and, where phonotactic restrictions allow it, in at least two different positions of the word. The general infrequent occurrence of the sound /ʒ/, however, limited the number of suitable words for young children so that this phoneme is only assessed four times as a singleton.
2) Bimorphemic contexts should be avoided as morpheme structure can influence children's phonological productions due to the frequent co-occurrence of SSD and developmental language disorder (Shriberg & Austin, 1998).
3) Each consonant occurs in at least two different vowel contexts to avoid an effect of a mutual influence of vowels and consonants on the children's productions (Stoel-Gammon & Dunn, 1985). An exception to this is /b/ which occurs with the vowel /ʌ/ only.
4) Each vowel phoneme occurs at least three times in the test and is assessed in stressed and unstressed positions.

5) The items vary in syllable structure and the number of syllables. Words with up to three syllables were included due to their diagnostic power (James et al., 2008b) and those containing consonant clusters were incorporated.
6) It was ensured that each phonological pattern that fulfils one of the following criteria could occur more than five times within the item set:
   a. The pattern is present in the typical speech acquisition of monolingual Turkish-speaking children.
   b. The pattern is frequently present in monolingual Turkish-speaking children with speech difficulties.
   c. The pattern was found to be frequently evident in the speech of typically developing Turkish-German bilingual children in previous pilot studies (Salgert, 2011; Ünsal & Fox, 2002).
7) As phonological competence is to some degree dependent on stored lexical information (Vogel Sosa & Bybee, 2011) and comorbidities of SSD and language impairments occur (Shriberg & Austin, 1998) an attempt was made to choose items that are well-known by young Turkish-German bilingual children. Since normative data on the vocabulary of Turkish-German bilingual children aged between 3;0 to 6;5 years did not exist at that time, children's books and nursery rhymes were used as a reference in the first place. In a second step, a work in progress version of the *Türkçe İfade ve Lisan Dizelges Araştırması* (TİLDA) [Turkish Language Instrument for the Early Assessment of Vocabulary Skills] (Sachse, Budde-Spengler, & Rinker, 2016) was used for reference. The TİLDA is a parent report form to collect data on the early vocabulary skills in Turkish of Turkish-German bilingual children in Germany. Sixty-nine percent of the items are consistent with items from the TİLDA and thus are regarded as age-appropriate for children up to the age of three. This included cultural specific words such as *cami* /d͜ʒʌːˈmi/, (mosque). The remaining 31% are added due to phonological criteria.
8) An effort was made to display the items with culturally sensitive ClipArt pictures to facilitate spontaneous naming. This included, for example, illustrations of culturally-specific tea glasses and coffee makers.

To confirm the suitability of the items and pictures for the target population and age group the newly compiled list of words was discussed with a Turkish-German bilingual SLT and then tested with four of the later test assistants in the project. Following their suggestions some ClipArt pictures were exchanged to facilitate recognition. This revised set was then discussed with two mothers of young Turkish-German bilingual children with SSD. According to their suggestions a few items were replaced by ones that were judged to be better known by young children but still fulfilled the linguistic criteria outlined above. Subsequently, the final set of items was piloted on one typically developing Turkish-

German bilingual child (5;6 years) as well as two children with speech and language impairment (4;2 and 9;3 years). The children could spontaneously name all the items except for those containing /ʒ/. These had to be elicited by delayed imitation. Since all words containing this phoneme had been fully replaced after the analyses of Salgert's (2011) pilot data and the number of child-appropriate words containing this sound is very limited it was decided that they remained in the assessment. This allowed for assessing the children's production of this Turkish phoneme in different word positions.

The final version of the TPA consists of 70 words – 61 nouns, two adjectives, two numeric adjectives, three colour adjectives, as well as two animal sounds. Twelve of the test items are monosyllabic, 46 disyllabic and 12 trisyllabic. A total of six words with syllable-final ($n = 1$) or word-final consonant clusters ($n = 5$) were included as well. To provide an engaging and child appropriate task, ClipArt pictures printed on A6 cards in colour were used to display the stimulus items. See Figure 3.4 for an example and Appendix G - Appendix L for the full list of Turkish test items and pictures.



No.	**Item**	**IPA**	**Prompts**	**Child's Realisation**	**Phonol. Patterms**
6	Cami	d͜ʒʌːˈmi	Ⓔ Ⓒ Ⓘ		

Note: **E**: explanation of the item's function, **C**: choice between two words, **I**: imitation

Figure 3.4: Example item from Turkish Phonology Assessment for Turkish German bilingual children: picture and section from record sheet

3.3.1.2 Administration of the single-word naming tests

In both, the PLAKSS-II and the TPA, participants were asked to spontaneously name the test items. The examiner prompted their productions by posing the questions *"What is this?"* or *"What do you see in this picture?"*. In the event of the item being unknown to the child the researcher explained the object's function. If this was not helpful to the child a choice of two words was given with the target word always being named first. This should prevent the occurrence of a direct repetition effect. Only in the case a child picked the wrong word from the choice and thus did not seem to know the word, he/she was asked to repeat the item. Both spontaneous and imitated response were included in the analyses (Goldstein, Fabiano, & Iglesias, 2004). Occasionally, children named the items in the non-target language. If this happened the examiners praised the child for their effort and then asked if they could say the word in the other language. Objectivity in administering the TPA was ensured by providing clear instructions for both the administration and recording of the test on the record form. This included specific eliciting questions to pose if "What is this?" was not applicable, for example, in case a verb or an adjective were requested.

There were no practise items included in the assessments since all words were supposed to be known by children of the included age bands and task demands were low as picture naming is a common activity in child care settings (Stackhouse & Wells, 1997).

At t1, the administration of the PLAKSS-II lasted for 27:23 minutes on average (range: 07:45 - 59:10), whereas the TPA, having 26 items less, took the children 24:16 minutes on average (range: 11:16 - 48:10) to complete. These times changed slightly for t2 when the children were 12 - 15.5 months older. Naming all test items took the children 18:39 minutes on average (range: 06:27 - 35:40) for the PLAKSS-II and 25:25 minutes on average (range: 12:55 - 48:59) for the TPA at t2. The differences were significant for German (Mdn_{t1} = 23:31, Mdn_{t2} = 17:38): $z = -5.34$, $p < .001$, $r = -.81$, but not for Turkish (Mdn_{t1} = 23:02, Mdn_{t2} = 23:58): $z = -1.00$, $p = .161$, $r = -.15$).

3.3.1.3 Psychometric evidence of the applied single-word naming tests when used with Turkish-German bilingual children

To investigate whether the phonology assessments administered in this project reliably examine what they are intended to examine in Turkish-German bilingual children, the tests' item difficulties (as measured by item familiarity) as well as inter- and intra-rater reliabilities were explored. These analyses were considered important for both assessments as (a) the PLAKSS-II's validity and reliability have not been reported for Turkish-German bilinguals and (b) since the TPA is a newly developed test tool that lacks psychometric evidence.

The first part of this subsection deals with the item familiarity of the assessments and the second one presents the outcome of the reliability analyses. Further, validity evidences, such as developmental trends, predictive utility and test 'norms' are reported in Chapter 4. Data on the concurrent validity of the TPA are not reported for two reasons. First, no additional test sessions with different Turkish phonology tools or expert judgements were included in this project as assessing the psychometric properties of the TPA was not the primary focus of the present research. Therefore, its diagnostic accuracy was also not investigated. Nevertheless, theoretically grounded rationale for the evaluation of children's performances as well as their categorisation are provided in Section 3.4.8. Second, the assessment tools available to comprehensively assess Turkish-German bilingual children's phonological skills in Turkish are not standardised and lack in linguistic completeness (cf. Section 2.1.3) so that comparison outcomes would have been difficult to interpret and most likely yielded to less meaningful results.

Children’s item familiarity on the PLAKSS-II

To identify how familiar Turkish-German bilingual children aged 3;0 - 6;5 years are with the items (vocabulary) in the PLAKSS-II, the percentage of children who could spontaneously name an item was calculated for each item and at both time-points. Subsequently, the items were allocated to one of ten percentage bands starting at 0-10% and ranging to 91 - 100%. The types of prompts that were provided to yield an elicitation of the target item when spontaneous naming was not possible were recorded as well and are presented per item in Appendix M and Appendix N.

As can be seen from Figure 3.5, allocations revealed that more than half (57%) of the items of the PLAKSS-II were named spontaneously by more than 50% of the participants at t1 (coloured in green in the chart). Within this range the largest number of items was named spontaneously by 71 - 80% of the children.

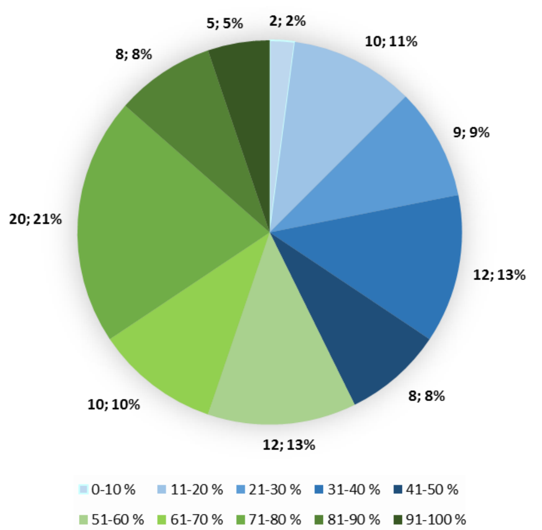


Figure 3.5: Participants’ familiarity with the items in the PLAKSS-II at t1

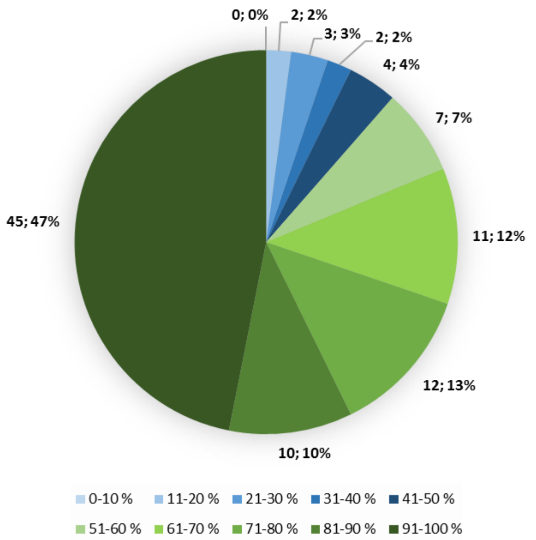


Figure 3.6: Participants’ familiarity with the items in the PLAKSS-II at t2

Note: Green coloured pie pieces refer to those items that could be spontaneously produced by more than 50% of the sample. Blue coloured pie pieces refer to those items that were spontaneously named by up to 50% of the sample.
Both pie charts display the actual values (number before the semicolon) and the percentages (number after the semicolon) of the items in the test per category.

The proportion of items being named spontaneously by more than 50% of the children significantly increased to 89% at t2 (see Figure 3.6). Here, the largest number of items was named spontaneously by 91 - 100% of the children. Further, all items at t2 could be spontaneously named by at least 11% of the children. It should be borne in mind, though, that sample sizes at t1 and t2 differed significantly.

Children's item familiarity on the TPA

As for the items in the PLAKSS-II, the percentage of children who could spontaneously name an item of the TPA was calculated for each item and at both time-points. Subsequently, the items were allocated to the ten different percentage bands. Figures with the types of prompts provided in case items could not be elicited are presented in Appendix O and Appendix P.

As Figure 3.7 shows, the items of the TPA were less well known by the children than the items of the PLAKSS-II. At t1 only 40% of the TPA-items were spontaneously named by more than 50% of the children (coloured green in the chart). Within this range the largest number of items was spontaneously named by 71 - 80% of the children.

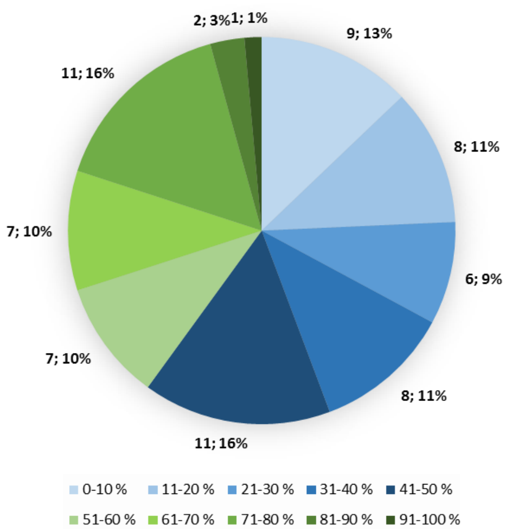


Figure 3.7: Participants' familiarity with the items in the TPA at t1

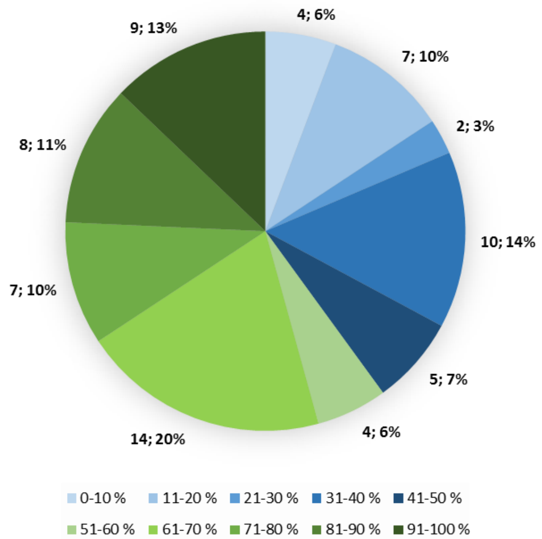


Figure 3.8: Participants' familiarity with the items in the TPA at t2

Note: Green coloured pie pieces refer to those items that could be spontaneously produced by more than 50% of the sample. Blue coloured pie pieces refer to those items that were spontaneously named by up to 50% of the sample.
Both pie charts display the actual values (number before the semicolon) and the percentages (number after the semicolon) of the items in the test per category.

However, as can be seen from Figure 3.8, the number of items that could be named spontaneously increased significantly at t2 to 60% (coloured green in the chart). The largest number of TPA items that did not require any prompt could be named by 61 - 70% of the children.

Inter-rater reliability

A randomly selected 10% of the German and Turkish t1- and t2-data were analysed regarding the agreement between examiners on broad phonetic transcriptions of children's productions and the analyses of children's phonetic-phonological patterns. First, data in both languages were transcribed and analysed by the author of this research project, a qualified SLT with research and extensive phonetic transcription experience. The German data were then re-transcribed by two qualified SLTs (one at each time point) with research and transcription experience. Data in Turkish (t1 and t2) were re-transcribed by a German SLT and PhD-student with extensive transcription experience. Agreements were calculated by point-by-point analyses for both vowels and consonants. No words were excluded from the analyses. In addition, a German qualified SLT with research experience re-analysed the phonological patterns in German and Turkish for both time-points of a randomly selected group of children (i.e., 10% of the sample). Qualified Turkish-German bilingual speakers were not available for these analyses.

As presented in Table 3.7, the point-by-point analyses at t1 revealed an average agreement of above 90% for transcriptions and above 85% for phonological pattern analysis in German and Turkish.

Table 3.7: Inter-rater reliability of single-word naming assessments at t1 and t2

Assessment	**Time**	**Transcription**		**Phonol. pattern analysis**	
		M	Range	*M*	Range
PLAKSS-II	t1	90.24%	81.97 - 95.80%	85.34%	77.22 - 94.74%
	t2	93.01%	91.77 - 93.64%	89.48%	79.57 - 95.36%
TPA	t1	90.00%	87.24 - 93.14%	86.29%	77.38 - 92.11%
	t2	92.76%	90.14 - 94.90%	87.43%	78.60 - 91.61%

At t2, the average agreement between transcribers was above 90% for transcriptions and above 85% for the analyses of phonological patterns in both languages. Across both test tools the average percentage of agreement increased from t1 to t2 when the children were at least 12 months older.

Intra-rater reliability

A further randomly selected 10% of the data were re-transcribed and re-analysed by the author to ensure intra-rater reliability. This included the re-transcription and re-analysis of all named items and included the transcription of phonetic variations (e.g., interdental realisations). The point-by-point analyses calculated revealed an agreement of above 90% for transcriptions and pattern analyses in both languages and at both time-points (see Table 3.8). Similar to the inter-rater reliability, the percentages for average agreements increased slightly between t1 and t2.

Table 3.8: Intra-rater reliability of single-word naming assessments at t1 and t2

Assessment	Time	Transcription		Phonol. pattern analysis	
		M	Range	*M*	Range
PLAKSS-II	t1	92.98%	85.71 - 97.46%	95.90%	89.70 - 100.00%
	t2	96.07%	94.30 - 98.10%	96.32%	92.98 - 99.06%
TPA	t1	91.39%	87.92 - 96.71%	94.98%	90.43 - 100.00%
	t2	93.14%	89.12 - 95.62%	92.38%	89.11 - 95.96%

3.3.2 Psycholinguistic output tasks at t2

The psycholinguistic tasks at t2 were aimed to (a) investigate children's ability to produce the phones of their languages in isolation and to assess their motor abilities (phone imitation), (b) to obtain data about the children's ability to accurately devise new motor programs without relying on their lexical representations (NWR), and (c) to find out whether children could generate newly assembled motor programs for unknown words repeatedly and consistently (consistency of NWR). In the following subsections, the individual psycholinguistic tasks are described and their reliabilities presented.

3.3.2.1 Phone imitation

Within this task, children were asked to repeat all German and Turkish speech sounds after the examiner. In case a child did not imitate a phone correctly, two further prompts were given: 1) the examiner reproduced the phone and asked the child to imitate it once more, 2) the child was asked to look at the examiner's mouth and its movements during the production of the phone and then asked to produce the phone one last time. It took the children 03:30 minutes on average to complete this task (range: 02:16 - 07:41).

3.3.2.2 Non-word repetition and consistency of non-word production

For the NWR and consistency of non-word production tasks (which were administered together) quasi-language-independent stimuli were used. This means, the items were constructed as language-independently as possible but contained a few language-specific features (see further below for a detailed description of the phonetic-phonological composition of the words). The use of quasi-language-independent stimuli was beneficial in two ways: 1) it was time-efficient as only one task needed to be administered (as opposed to one for German and one for Turkish), and 2) more importantly, this allowed for tailoring the stimuli specifically to the language combination of Turkish and German. This was important since the performance in NWR-tasks can be influenced by language knowledge so that the representativeness of results elicited via monolingual versions was questioned (see Sections 2.2.4.2 and 2.2.4.3).

The stimuli used in this project were taken from a 40 item quasi-language-independent NWR-task designed within the COST Action IS0804 by Engel de Abreu and dos Santos (in prep.). The items included only a small set of phonemes which are shared across a large number of languages (i.e., /p, k, t, f, l, ʃ/ and /u, a, i, ɑ/), included consonant clusters in various positions of the word and were allocated a Luxembourgish stress pattern, the language in which Engel de Abreu and dos Santos' (in prep.) version was applied in. Since not all features of the task's items were compatible with the German and Turkish sound systems a subset of 14 two- to five-syllabic items was selected that would meet the following criteria (see Table 3.9):

- Not containing any initial consonant clusters. This was to ensure that all items were in line with the phonotactic restrictions of both languages, Turkish and German.
- An even number of items with two- to five-syllables (three each) incorporating various syllable structures (i.e., combinations of CV and CVC) should be included to (a) control for a length effect and (b) to make use of polysyllabic word's discriminant power (James et al., 2008b).
- Two disyllabic non-words should be selected as practice items.

Table 3.9: Items for NWR- and consistency task

2-syllabic	3-syllabic	4-syllabic	5-syllabic
[ˈpilu]*	[kiˈfa:pu]	[pufiʃˈku:fi]	[kiʃuˈpi:faku]
[ˈfulpi]*	[kuˈpalfi]	[kalfiˈpa:ku]	[fapuˈka:ʃifal]
[ˈfa:ku]	[puˈfa:ki]	[ʃufiˈka:pi]	[puʃiˈfa:kilpa]
[ˈfapuʃ]			
[ˈpilfu]			

Note: *****practice items

Since the selected items of the NWR-task originally included the vowel /ɑ/, which is neither part of the German nor the Turkish phoneme inventory, it was substituted by the vowel phoneme /a/. This sound is part of the German phoneme inventory but is only recognised as a phoneme of standard Turkish by some authors (Topbaş, 1997). It was nevertheless included as Turkish-German bilinguals have been found to frequently produce this sound for /ʌ/ in their Turkish productions (Salgert, Fricke, & Wells, in prep.).

The author pre-recorded herself speaking the items using a ZOOM H4n digital audio recorder. Recording noise was cleared using RX4 Audio Editor (iZotope, 2014). Subsequently, the stimuli were incorporated into a PowerPoint presentation. The NWR- and consistency task were administered as a teddy-game in which the teddy was aiming to receive a pot of honey for which he needed help by the child (i.e., by repeating the non-words). The task started with two practice items to familiarise the children with the task (see instructions and record form in Appendix Q and Appendix R, respectively). Each stimulus was played to the child one at a time at a comfortable listening level via an Acer Aspire M Laptop loudspeaker in a quiet room of the children's nursery. The volume was adjusted to the individual child's need. Participants were asked to repeat each item once for the NWR-task and immediately after this three additional times for the consistency task. Hence, a stimulus was repeated consecutively four times in total. In case the child did not immediately repeat the word presented the examiner tried to trigger them with: *"Could you repeat what the teddy has just said? And could you say the word another three times?"* All items were represented by a coloured box in the PowerPoint presentation (i.e., blue for practice items and yellow for test items see Figure 3.9). After the completion of an item its colour faded, thus indicating how many items were still to be done. Following the completion of all items the tester presented the next slide which showed the teddy having received the pot of honey.

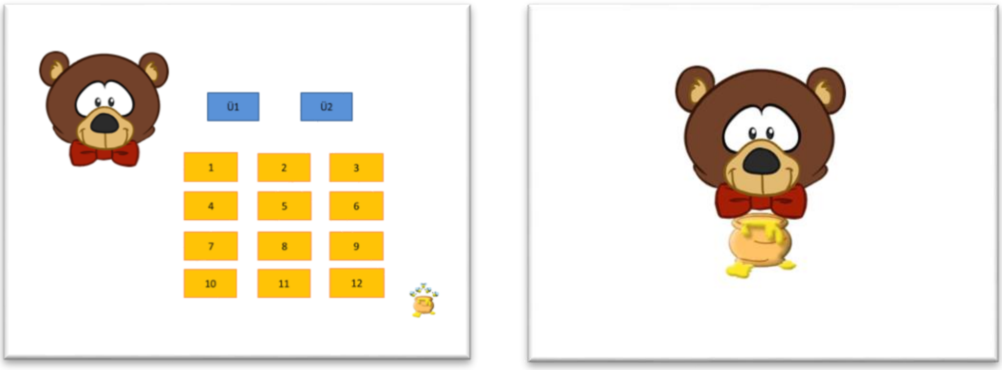


Figure 3.9: Presentation of NWR- and consistency task at t2

Throughout the task the stimuli were allowed to be replayed to the child only once in order to prevent a learning effect (Stackhouse et al., 2007). This was considered in three occasions: 1) if the child started talking while the audio file was played so that the stimulus was unlikely to be heard by the child in full, 2) if the child was distracted (e.g., by an external noise) or 3) if the child requested a replay.

The NWR- and consistency task took the children on average 05:22 minutes (range: 04:14 - 07:43) to complete.

3.3.2.3 Reliabilities of the psycholinguistic tasks

To assess the reliabilities of the quasi-language-independent psycholinguistic tasks calculations of internal consistency (Cronbach's alpha) as well as of inter- and intra-rater agreement on transcriptions were computed. According to Nunnally (1978), test reliability scores of $\alpha = .70$ or above are regarded as acceptable if a test is used in preliminary research. For basic and applied research Cronbach's alpha coefficients of ≥ .80 and ≥ .90, respectively, are minimally acceptable reliability levels. Inter- and intra-rater transcription reliabilities for the three tasks were carried out in the same way as for the single-word naming tasks (see Section 3.3.1.3) and included transcripts of randomly selected participants ($n = 5$, 11.6% of the sample).

Table 3.10 indicates that the internal consistency of the NWR-task was acceptable for preliminary research and approaching acceptable reliability levels for basic research. In contrast, the reliabilities of the consistency and phone imitation task were just below the acceptable value for a good reliability estimate of preliminary research.

Table 3.10: Reliabilities of applied psycholinguistic tasks

Task	*α*	Inter-rater		Intra-rater	
		M	Range	*M*	Range
Phone imitation	.662	95.50%	92.50 - 100.00%	91.32%	83.45 - 97.67%
NWR	.785	91.97%	85.00 - 97.05%	95.64%	93.27 - 98.00%
Consistency	.686	93.21%	87.87 - 95.11%	94.22%	91.94 - 97.35%

Note: **α**: Cronbach's alpha

Transcription agreement for all tasks in both conditions (i.e., inter- and intra-rater agreement analysis) was above 90%, overall ranging from 83.45 to 100%, and thus an indicator of good inter- and intra-rater agreement.

3.4 Scoring

The children's productions in both Turkish and German were analysed separately for each participant. Due to the scope of this thesis, the focus was on children's phonological realisation of consonants, thus, phonetic analyses and those including vowels are generally not reported. All exceptions to this are explained in the descriptions of the respective analyses and their corresponding scoring criteria.

The analyses conducted at t1 and t2 included compiling consonant and consonant cluster inventories, calculating the percentage of consonants correct – adjusted (PCC-A) and analysing phonological variations (PhonVar) as well as infrequent variants (InfVar). The psycholinguistic tasks at t2 were analysed regarding the number of correctly imitated phones, the number of accurately repeated non-words and the consistency of non-word productions. Scoring procedures for all analyses are described in detail in the following.

Standard pronunciation for comparison

Dialectal/accentual variations from the standard pronunciation of German and Turkish were considered at all times and not scored as an error. Further, to avoid misidentification of speech errors, all types of CLT (phonetic and phonological) that have been reported to frequently occur in adult or child bilingual speakers of Turkish-German were taken into account (cf. Chilla et al., 2010; Keim, 2007; Salgert et al., 2012; Tekin, 2011; Tekin & Colliander, 2010). This was extended to any pronunciation changes to standard Turkish observed in the speech of the Turkish-German bilingual test assistants. It was noticed, for example, that none of them consistently used the Turkish allophones for /k, g, l, v/ in the phonotactic contexts they were required – they usually substituted /c, ɟ, ɫ, ʋ/ by [k, g, l, v]. Thus, this was accepted in the children's speech as well. A full list of the realisations scored as phonemically correct for each item on the single-word naming assessments is attached in Appendix S and Appendix T. This includes CLT phenomena.

3.4.1 Consonant inventories

To obtain information on the children's (phonemic) acquisition of consonants in each language, consonant inventories of Turkish and German were generated. A consonant was considered to be acquired by an individual if it occurred in at least two thirds of the possible occurrences in the speech sample collected (Fox & Dodd, 2001; Ray, 2002; Topbaş & Yavaş, 2006; Yang & Zhu, 2010). A consonant was considered to be acquired by an age group if at least 75% of the children in the age group had acquired the sound to the criterion described above. A consonant was considered to be mastered by an age group if at least 90% of the children in the age group had acquired the sound to the criterion described above. This is in line with a number of several other studies (e.g., Dodd, 1995; Fox & Dodd, 1999; Topbaş & Yavaş, 2006; Zhu, 2000).

In addition to phonetic and dialectal variations the following aspects were considered in the scoring process:

- Lateralisation of alveolar or postalveolar sibilants was generally considered to be a phonetic distortion of the target sound and not scored as an error. In case both alveolar and postalveolar sibilants were substituted by /ɬ/ or /ɮ/, though, a correct phonological representation of /s, z, ʃ, ʒ, ç/ could not be assumed. The consonant(s) was/were then not evaluated as correct.
- Some children produced German and Turkish postalveolar fricatives as (alveolo)-palatal equivalents (e.g., /ʃ/→ [sʲ/ɕ]). This was scored as a phonetic variation of postalveolar sibilants if only these sounds were substituted (Bernhardt, Romonath, & Stemberger, 2015). In case a child also substituted the alveolar sibilants by these sounds, a correct phonological representation of /s, z, ʃ, ʒ/ could not be assumed and the sound(s) was/were scored as phonemically incorrect.
- Deaffrications of word-initial /p͜f/ in the items *Pflaster* (plaster) /ˈp͜flastɐ/ and *Pferd* (horse) /p͜fɛɐt/ was not scored as incorrect. This speech pattern is frequently found in German adult speakers and considered to be a variation of the High Standard German pronunciation (Fox, 2011). Moreover, only 11 participants produced this affricate correctly in *Pflaster* and 20 children in the item *Pferd*. More than 50% of the sample pronounced the word-initial /p͜f/ as [f] ($n = 56$ (67%) for *Pflaster* and $n = 61$ (72.6%) for *Pferd*).
- Children sometimes added a /n/ to words ending on a Schwa in German which converted them into a verb or the plural form. [+n] is the flexive morpheme to mark a plural in trochaic words ending on a Schwa. But [+ən] is also the suffix to mark the infinitive of a verb, thus, the word structure the children produced could be either or (e.g., *Rutsche* (slide) /ˈʁʊtʃə/ → *R/rutschen* (slides/to slide) [ˈʁʊtʃən]). Since the additional phoneme /n/ is in line with phonotactic restrictions of German and represents a word with linked semantics the supplemented /n/ was ignored for the analysis of the child's consonant inventory as it is not part of the target item.
- In case a word-initial weak syllable was deleted the following syllable-initial consonant was counted for the inventory as syllable-initial within word. In this case, it was assumed that the children's underlying representations of the words are correct because weak syllable deletion tokens were low in the sample (e.g., the /b/ in /gəˈbʊɐt͜stak/ was scored as syllable-initial within word correct if the weak syllable /gə/ was deleted). The same was applied to syllable-final consonants within word when the following weak syllable was not realised.

3.4.2 Consonant cluster inventories

Consonant clusters (CCs) are a frequent syllable structure in German, whereas in Turkish they are the exception. Only a few word-final CCs exist (Kopkallı-Yavuz, 2010; Kornfilt, 2011; also see Section 1.1.2). Against this background, assessing the children's acquisition of CCs and their strategy to deal with this frequent syllable structure in German was of great interest. Since most German CCs assessed within the PLAKSS-II are tested only once, solely the first occurrence of a cluster in the test was evaluated. Hence, in case a cluster was assessed more than once but a child did not name the CC-item which occurred first, the next item assessing this cluster which was realised by the child was scored. This ensured that all clusters were evaluated in the same way, both within and across assessments (i.e., as in the TPA).
Separate CC inventories for each child and their two languages were compiled. German cluster inventories were further divided into syllable-initial and syllable-final inventories.

In general, a cluster had to be produced at least once to be considered acquired by a child. However, three different criteria (one at a time) were applied to score a cluster as correct:

1) If it was produced phonemically correct. This scoring was consistent with the one for consonants.
2) If the cluster elements were produced phonemically correct but split up by an epenthetic vowel/schwa (e.g., *Blume* (flower) /ˈbluːmə/ → [bəˈluːmə]). This criteria was used because vowel epentheses in clusters is a frequently reported phenomenon in bilingual children (e.g., Preston & Seki, 2011), has been reported to be a typical interactional pattern between German and Turkish and reflects the phonotactic rule for loan words containing initial clusters in Turkish (Kornfilt, 2011).
3) If its syllable shape was produced correctly in which case the number of cluster elements was accurate independently from their phonemic correctness (e.g., *Schmetterling* (butterfly) /ˈ<u>ʃm</u>ɜtɐlɪŋ/ was scored as correct if it was realised as [ˈ<u>sm</u>ɜtɐlɪŋ], [ˈ<u>ʃp</u>ɜtɐlɪŋ] etc., but not if it was produced as [ˈ<u>s</u>ɜtɐlɪŋ], [ˈ<u>ʃpl</u>ɜtɐlɪŋ] etc.). This criterion was considered to be important as it may provide insight into whether children have understood/acquired the concept of consonant clusters at all.

To analyse the acquisition of a cluster by a certain age group the same 75%- and 90%-criteria as for the consonant inventories were applied for all three scoring criteria outlined above.

3.4.3 Percentage of consonants correct - adjusted

For comparability purposes with other studies the children's PCC-A-scores were calculated. The PCC-A was preferred over the standard PCC (Shriberg & Kwiatkowsky, 1982b) since it accepts all *"common clinical consonant distortions"* in the children's productions as correct (Shriberg et al., 1997a, p. 711). Since the present research focuses on the acquisition of phonology this definition was extended to all phonetic variations children produced and thus allowed for accounting for dialectal influences and cross-linguistic transfer. In general, the calculations were run following the guidelines in Shriberg et al. (1997a). Instead of using conversational speech samples, however, the measurement was applied to the present single-word naming data. The results were thus not applicable for a classification into the four severity categories proposed in the validation study of the PCC (Shriberg et al., 1997a). Nevertheless, they allowed descriptive within subjects and across language comparisons.

3.4.4 Phonological variations

In order to quantitatively and qualitatively describe and understand Turkish-German bilingual children's developmental realisations of single words, their PhonVar from adult-like speech were evaluated. First, the overall number of phonologically variated productions were counted per child. Second, these variations were differentiated into *infrequent variants* and *phonological patterns* depending on their frequency of occurrence, since a single instance of a variation may be due to developmental fluctuation or occur by chance. A frequent occurrence of a phonological variation, however, demonstrates a certain tendency in a child's speech (Dodd et al., 2003).

For the present research project, the number of PhonVar was restricted to variations affecting consonants. Due to this, the measure is similar to the PCC-A-measure as both describe the degree of inaccurate speech just from a different perspective. Whereas PCC-A functions to measure correctness, PhonVar are used to measure deviations (Dodd, 1995, 2005; Shriberg & Kwiatkowsky, 1982b). A child who presents with severe speech difficulties will probably show a large number of PhonVar and only achieve a low PCC-A-score. However, the number of PhonVar is not the reverse count of the number of consonants correct (converted into the PCC-A-score) since a single consonant can be affected by more than one phonological pattern (e.g., *Schiff* (ship) /ʃɪf/ → [pɪf], in which the /ʃ/ is stopped and additionally assimilated to the place of articulation of /f/). Therefore, both measures were applied.

If a type of phonological variation occurred up to four times, it was considered to be infrequent and labelled as an *infrequent variant (InfVar)*. These less frequent production patterns are usually not considered in phonological pattern analyses in the literature, compare for example, Dodd, Holm and Li (1997). However, as they might shed light on the variability and consistency in the children's productions they were considered in Fox (2016) as well as in the present research project. The number of InfVar per participant and language was counted.

All types of PhonVar occurring five or more times were considered to be frequent and were labelled as *phonological patterns*. The number of five necessary occurrences is fairly arbitrary and still needs to be supported empirically (McReynolds & Elbert, 1981), however, this measure has been frequently applied in previous research (Dodd et al., 2003; Holm & Dodd, 1999a, 1999c, 2006; Holm et al., 1997) and is recommended to be applied to the PLAKSS-II data (Fox-Boyer, 2014c). Thus, to ease comparisons across studies, follow the recommended procedures for the PLAKSS-II and to keep scoring procedures consistent across the two single-word naming assessments, five instances of occurrences of a phonological pattern are used as a cut-off score in this study. Due to the small number of participants per age group, a pattern was regarded as being present in a certain age group if at least 15% of the children in an age group exhibited this pattern. However, those patterns which occurred in only 10 - 14% of the children in an age group were also reported for comparison purposes with previous research (Dodd et al., 2003; Fox & Dodd, 1999; Holm, 1998; Law & So, 2006; Zhu, 2000).

Variations from adult-like speech that were of phonetic nature (e.g., a lisp) were not considered apart from those that were regarded to be caused by a cross-linguistic interaction between Turkish and German. These were identified by a contrastive approach towards German and Turkish phone inventories and phonotactic restrictions and listed separately as part of CLT phenomena.

3.4.5 Phone imitation

Children were considered to be able to articulate a phone of their languages if they could imitate it correctly once in isolation (cf. Dodd et al., 2002). All non-dialectal phonetic variations of consonants (e.g., the interdental realisation of the alveolar fricatives /s/ and /z/) were scored as incorrect. Similar to Williams and Stackhouse (2000) the number of correctly imitated phones was computed per child and formed the basis for age group data. Since shared phones were only assessed once, the maximum score for this task was 26.

3.4.6 Accuracy of non-word repetition

Children’s repetitions of non-words were scored regarding their phonemic accuracy in comparison to the target stimulus. Only fully phonologically accurate repetitions were scored as correct. This is a time-efficient way of scoring that is in line with previous research (dos Santos & Ferré, 2016; Nathan, Stackhouse, Goulandris, & Snowling, 2004a; Vance et al., 2005) and has especially been found for quasi-language-independent stimuli to be equally sensitive when identifying children with language impairment as using the measure of PCC (Boerma et al., 2015). The number of accurately produced repetitions of the non-words was calculated for each child and could result in a total score of 12. Practice items were excluded from scoring and analysis.

3.4.7 Consistency of non-word production

All four repetitions of each non-word (i.e., including the one during the NWR-task) were included in the analysis. The child gained a score of one per two consistent repetitions (i.e., two identical realisations irrespective of whether they were accurate compared to the stimulus or not) which could result in a maximum score of 36. This is somewhat different to the scoring systems used in previous studies in which usually one score was given if all repetitions of an item were identical (e.g., Holm, Crosbie, & Dodd, 2007). The scoring used in the present work, however, allowed for scoring fine differences in children’s degrees of consistency (i.e., by making a difference if three out of four repetitions were identical or only two).

As for all previously described measures, phonetic variations of consonants and vowels were not scored as inconsistent. In addition, if vowels in weak syllables were shortened (as is common in German where also the vowel quality of /i, o, u, e/ changes to [ɪ, ɔ, ʊ, ə] respectively; Fagan, 2009) these were not scored as an inconsistent production, whereas a change in both height and rounding of the vowel would be scored as an inconsistent production. Also, stress pattern differences across repetitions were not scored as a difference since the focus of this task was just phonemic accuracy. See Table 3.11 for four examples of what type of realisations have been scored as consistent.

Table 3.11: Example responses in the consistency task which were scored as consistent

Target	1st Repetition	2nd Repetition	3rd Repetition
ʃufi'ka:ti	ʃufi'ka:ti	ʃufi'kati	ʃufi'ka:ti
'fulpi	'fulpi	'fʊlpi	'fʊlpi
ki'fa:po	ki'fa:po	ci'fa:po	ci'fa:po
kiʃu'pi:faku	bɪ'sa:fo	'bɪsa:fo	bɪ'sa:fo

As for the NWR-task, realisations of practice items were excluded from scoring and analysis.

3.4.8 Categorisation procedures at t1 and t2

In line with a large number of studies on bilingual children (e.g., Dodd et al., 1997; Fox & Dodd, 2001; Holm & Dodd, 1999a, 1999b, 2006; Holm et al., 1997; Holm et al., 1999; Lee et al., 2015a; Preston & Seki, 2011), hypotheses about the quality of children's phonological acquisition were first postulated based on the nature of their phonological patterns at both time-points. Thus, a definition of the nature of the pattern is provided first and followed by a categorisation of the children's phonological performances.

3.4.8.1 Categorisation of phonological patterns

Each phonological pattern was categorised as either, *age-appropriate, delayed* or *deviant* compared to the Turkish-German bilingual cohort in this research as well as compared to existing data on monolingual children of each language. In line with the literature, age-appropriate phonological patterns are considered to reflect children's typical developmental strategies to deal with the acquisition of their complex phonological systems (Ingram, 1989b; Stoel-Gammon & Dunn, 1985). In contrast, delayed phonological patterns are interpreted as the child's difficulty in making the next developmental step towards correct pronunciation in time (Dodd, 1995). Children who produced deviant patterns were considered to use unique developmental paths and strategies to deal with the acquisition of a complex phonological system (Ingram, 1989b).

In the following the criteria to define the nature of the patterns used in this research are presented (cf. also Figure 3.10) and subsequently rationale for the decisions made are provided.

A pattern was considered to be *age-appropriate* if

- it was produced by 15% or more of the Turkish-German bilingual children in this age group
  OR
- it has been reported to be age-appropriate for monolingual German- or Turkish-speaking children.

A pattern was considered to be *delayed* if

- it occurred in less than 15% of the Turkish-German bilingual children in an age group but was evident in at least 15% of the children in an at least six months' younger age group
  OR
- it occurred in less than 15% of any age group in the Turkish-German bilingual cohort but was considered to be delayed (at least six months) for this age group in monolingual children of both languages (if applicable).

A pattern was considered to be *deviant* if

- it was not produced by at least 15% of the children in any age group in the Turkish-German bilingual cohort
  AND
- it was also not reported to be typical (either age-appropriate or delayed) for monolingual children of both languages.

These definitions are similar to previous work on monolingual and bilingual children (e.g., Dodd et al., 2003; Fox & Dodd, 1999, 2001; Holm, 1998; Law & So, 2006; Lee et al., 2015a; Topbaş, 2006; Zhu, 2000) in that they use a certain percentage of children in the same or a younger age group of a 'normative' sample for comparison. However, in previous studies a cut-off value of 10% for the patterns was used instead of 15%. Due to the small sample size in the present research, a 10% cut-off was regarded inappropriate due to the risk of identifying false-positive patterns (i.e., categorising a pattern as typical for normal development which may actually be deviant). Thus, the cut-off was elevated to 15% (see Section 3.4.4).

Monolingual data – which are based on large normative samples – were used for additional reference since even a cut-off value of 15% could be misleading in the present small sample and it was intended to make decisions in the child's best interest. Thus, patterns which were not present in at least 15% of the bilingual cohort but were considered typical for either German or Turkish monolinguals, were scored to be typical for Turkish-German bilinguals (see Figure 3.10).

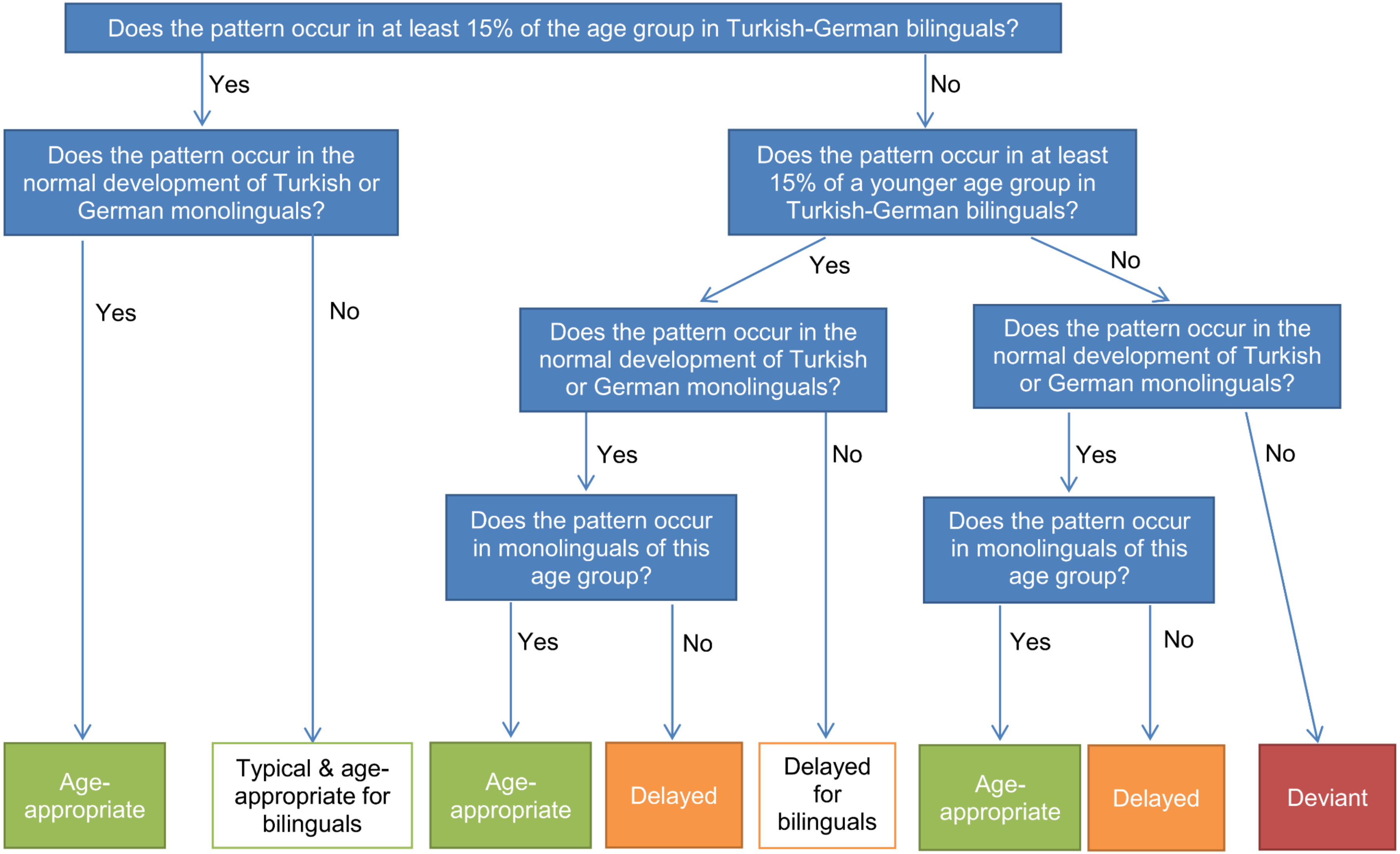


Figure 3.10: Decision tree to categorise a phonological pattern as age-appropriate, delayed or atypical for Turkish-German bilingual children

3.4.8.2 Categorisation of children's performances

Current research suggests that a SSD will present itself in all of a child's languages as speech output of bilinguals is assumed to be generated, restricted and managed by one single underlying mechanism (Dodd et al., 1997; Hambly et al., 2013). Thus, children's performances in both Turkish and German had to demonstrate some evidence for an atypical development in order to be labelled as atypical.

Besides the nature of children's phonological patterns in Turkish and German, their number of InfVar in each language was considered for the categorisation process, since these were believed to provide valuable insight into children's variability and consistency of phonological production. Additionally, the number of InfVar would add a quantitative aspect to the otherwise only qualitative analysis. For this, numbers higher than 1.25 *SD* above the age group mean (i.e., more InfVar) were considered to be weak performances, whereas numbers smaller than or exactly at 1.25 *SD* above the age group mean were considered to be normal/good performances. This cut-off was selected for two reasons: First, scores below (or above, depending on the measure) 1 *SD* of the age group mean have been frequently used to identify language difficulties (McLeod & Harrison, 2009; Reilly et al., 2014). Second, a study by Fox and Dodd (2001) revealed that children with SSD perform between 1 and 2 *SD* above the age group mean of typically developing children on the measure of percentage of incorrect phonemes. Thus, based on the nature of their phonological patterns and their number of InfVar in both languages children's overall phonological performances were categorised as follows:

<u>Typical development</u>

Children fell into the category of typical development if they did not produce any phonological patterns at all or only showed age-appropriate phonological patterns for either monolinguals and/or Turkish-German bilingual children in both languages. Additionally, children's number of InfVar should not exceed 1.25 *SD* above the age group mean in both languages (see Figure 3.11).

<u>Atypical development</u>

Children with the following phonological profiles were categorised as atypically developing:

- At least one deviant pattern in both languages
- At least one delayed pattern in both languages

- At least one deviant pattern in one language and at least one delayed pattern in the other language
- At least one deviant pattern in one language but age-appropriate or no patterns in the other languages but a high number of InfVar (i.e., above 1.25 *SD* above the age group mean) in both languages
- At least one delayed pattern in one language but age-appropriate or no patterns in the other language and a high number of InfVar in both languages
- No clear patterns in both languages but a high number of InfVar in both languages.

No category

Children whose phonological profiles did not meet the criteria for either typical or atypical development were considered to be unable to be categorised. Their profiles were characteristic of very different performances in their languages (e.g., they only showed age-appropriate phonological patterns in language A but at least one deviant pattern in language B) and/or demonstrated a high number of InfVar (i.e., exceeding 1.25 *SD* above the age group mean) in only one language.

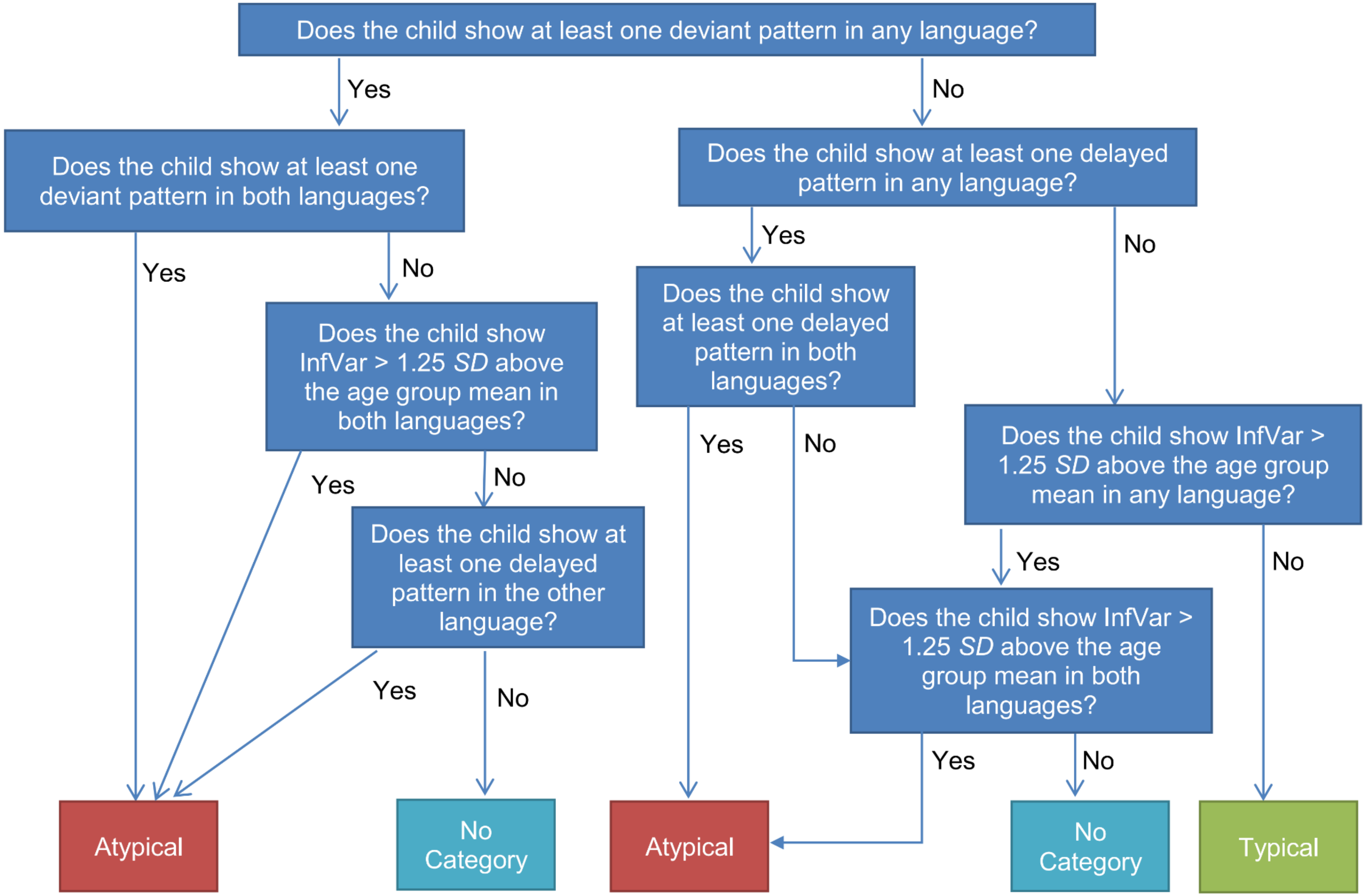


Figure 3.11: Decision tree to categorise the children's performances in the single-word naming assessments at t1

In addition to the measures used at both time-points, children's performances in the NWR- and consistency-task were considered in the categorisation process at t2. Those additional psycholinguistic data were hypothesised to provide insight into the nature of children's speech sound development and to help allocating especially those children who were unable to be categorised to either the typical or atypical category at t2.
To ensure consistency across measures, NWR- and consistency performances were also scored using a cut-off of 1.25 *SD* below the age group mean. This cut-off was further decided upon because children had already been categorised based on their phonological performance in the single-word naming tasks (pattern analysis and number of InfVar) so that the additional measures were intended to provide a further tendency regarding the nature of children's speech sound skills.

3.4.9 Missing data

There were a few children who refused to name all items of the PLAKSS-II and TPA or refused to imitate all Turkish consonants at both time-points (see Table 3.12). The percentage of missing data, however, was very small compared to the overall number of possible productions in the respective measures. Although missing data influenced the scores achieved by an individual, they only rarely affected age group results (e.g., the acquisition/mastery of a phoneme by a certain age group) and never changed the categorisation of a child's performances.

Table 3.12: Overview of missing data at t1 and t2

Measure	Language	Time	Children (*n*)	% of possible productions affected	Age group data affected?
Consonant inventory	German	t1	4	0.20	No
		t2	1	0.12	No
	Turkish	t1	12	0.26	No
		t2	7	0.37	Yes (acquisition of /v/ in 5;6 - 5;11-year-olds)
CC inventory	German	t1	4	0.31 (onset CC) 0.13 (coda CC)	No
		t2	1	0.26 (coda CC)	No
	Turkish	t1	0	0	No
		t2	1	0.39 (coda CC)	No
Phone imitation	German	t2	0	0	No
	Turkish	t2	4	1.51	Yes (*M* and *SD*, but categorisation of children remained unaffected)

Regarding phonological pattern analyses, there was always a large enough number of other opportunities for the pattern to occur so that the missing productions did not seem to have an effect on them either. All instances in which age group data were affected by missing data are indicated in the results chapters. To account for an effect on individual results, children's scores in the particular measures were calculated based on the actually attempted productions (cf. Vance et al., 2005). For PCC-A-scores, for instance, the number of correctly produced consonants was divided by the maximum number of consonants occurring in the items named by a child.

3.5 Procedure

At t1 and t2, children were individually assessed in a quiet room of their nurseries. To eliminate fatigue effects and a lack in attention as well as to keep the amount of code-switching to a minimum, the Turkish and German test sessions were scheduled for separate days (cf. Lleó & Kehoe, 2002). The time distance between the two sessions was dependent on the availability of the test assistants, the nurseries' schedules and the child's nursery attendance times. On average, test sessions were 12.6 days apart at t1 (range: 1 - 25 days) and 14.7 days apart at t2 (range: 3 - 27).

All assessments were administered by native speakers of each language. Assessments that were instructed in German (i.e., PLAKSS-II, German phone imitation, NWR- and consistency task) were administered by the author, a qualified SLT with research experience. Five Turkish-German bilingual SLT or linguistics students administered the Turkish test sessions (i.e., TPA and Turkish phone imitation). Prior to the data collection a comprehensive information and training session for all bilingual test assistants was conducted. In this two to three hours face-to-face meeting, the PhD project and its aims were explained and information on the general study procedure were provided. The test assistants had the opportunity to familiarise themselves with the assessment material and were subsequently trained in detail on its administration. Turkish test sessions were supervised by the author. It was further ensured that the order of German and Turkish sessions was counterbalanced. The allocation of the children to each of the two 'language-first-groups' depended on the test assistants' availability and the nurseries' schedules.

Incorrect realisations of the children were taken note of on-line on the respective record forms by all examiners. Additional audio recordings with a ZOOM H4n digital audio recorder ensured subsequent exact and reliable analyses. The speech data were then transcribed phonetically in IPA by the author while primarily referring to the audio recordings. Notes on the record forms were consulted in cases of auditory ambiguity. The author then analysed all data according to the scoring criteria outlined for each task.

4 RESULTS I – Phonological acquisition in Turkish-German bilingual children

The results of the single-word naming assessments conducted at t1 were analysed and are presented to address the first research question and its two sub-questions outlined in Section 2.3. This results chapter has been divided into four subsections:

1) Acquisition of consonants and PCC-A
2) Acquisition of consonant clusters
3) PhonVar from adult-like speech
4) Influential factors on phonological performance.

Results in this chapter are based on the analyses of the single-word naming tests and are presented separately for each language with those for German preceding those for Turkish. However, combined presentations are applied whenever direct comparisons are appropriate.

4.1 Acquisition of consonants and PCC-A at t1

To address research question 1a consonant inventories were compiled per participant and analysed separately for German and Turkish. Additionally, percentages of consonants correct – adjusted (PCC-A) were calculated for both languages (see Sections 3.4.1 and 3.4.3 for detailed descriptions of the scoring criteria applied).

4.1.1 Acquisition of German consonants

Children in the youngest age group (i.e., 3;0 - 3;5 years) had already acquired (75%-criterion) the following twelve German consonants: /m, n, b, p, d, t, k, f, s, x, l, j/ (see Table 4.1). Six of these phonemes were also considered to be mastered (90%-criterion) at this age: /m, n, b, d, k, j/. The older the children were the more consonants they had acquired so that by the age of 5;0 - 5;5 years children were only missing three consonants of their inventory (i.e., /ʃ, ç, t͜s/) and demonstrated mastery of two thirds of German consonants.

Table 4.1: Acquisition and mastery of German consonants at t1

Age group	3;0 - 3;5	3;6 - 3;11	4;0 - 4;5	4;6 - 4;11	5;0 - 5;5
Cons. / *N*	12	21	17	27	7
m	mastered	mastered	mastered	mastered	mastered
n	mastered	mastered	mastered	mastered	mastered
ŋ	not acquired	mastered	acquired	mastered	mastered
p	acquired	mastered	mastered	mastered	mastered
b	mastered	mastered	mastered	mastered	mastered
t	acquired	mastered	mastered	mastered	mastered
d	mastered	mastered	mastered	mastered	mastered
k	mastered	mastered	mastered	mastered	mastered
g	not acquired	not acquired	not acquired	acquired	mastered
f	acquired	mastered	mastered	mastered	mastered
v	not acquired	not acquired	not acquired	mastered	acquired
s	acquired	mastered	acquired	mastered	mastered
z	not acquired	mastered	acquired	mastered	acquired
ʃ	not acquired	not acquired	not acquired	not acquired	not acquired
ç	not acquired	not acquired	not acquired	not acquired	not acquired
x	acquired	acquired	acquired	mastered	mastered
ʁ	not acquired	not acquired	not acquired	acquired	acquired
h	not acquired	mastered	mastered	mastered	mastered
l	acquired	mastered	mastered	mastered	mastered
j	mastered	mastered	mastered	mastered	mastered
p͜f[1]	not acquired	not acquired	not acquired	acquired	mastered
t͜s	not acquired	not acquired	not acquired	not acquired	not acquired

Note: **Cons.**: consonant, **shared consonants with Turkish**, consonant mastered by age group, consonant acquired by age group, consonant not acquired by age group

[1]In word-initial position [f] was accepted as phonemically correct for /p͜f/ since this is the variation of German spoken in the North of Germany, where data have been collected.

This developmental trend in consonant acquisition was supported statistically by a moderate correlation[9] of the children's age with the number of acquired consonants, $r_s = .506$, $p < .001$. Consonants which were not acquired or not mastered by each age group mainly belonged to the sound classes of fricatives and affricates. The phonemes /g, ŋ/ (not acquired) and /p, t, l/ (not mastered) form exceptions. Further, the majority of consonants not yet acquired were comprised of language-specific sounds. Sounds which are shared between Turkish and German (i.e., /m, n, p, b, t, d, k, g, f, v, s, z, ʃ, h, l, j/) were rarely missing from children's consonant inventories (cf. Section 4.1.3).

[9] Throughout this thesis, the strength of correlations was interpreted as follows: r_s = .00 - .19 (negligible), r_s = .20 - .39 (weak), r_s = .40 - .59 (moderate), r_s = .60 - .79 (strong), r_s = .80 - 1.0 (very strong) in line with Evans (1996).

4.1.2 Acquisition of Turkish consonants

The acquisition and mastery of Turkish consonants by the individual age groups at t1 are presented in Table 4.2. Nearly half of the Turkish consonants were already acquired by the youngest age group (i.e., 3;0 - 3;5 years). These included: /m, p, b, t, d, k, f, s/ and /j/. Four of these consonants were already mastered at this age: /m, p, b, k/. The acquisition of Turkish consonants followed a developmental trend demonstrating that the older the children were the more Turkish consonants they had acquired, $r_s = .429$, $p < .001$.

Table 4.2: Acquisition and mastery of Turkish consonants at t1

Age group	3;0 - 3;5	3;6 - 3;11	4;0 - 4;5	4;6 - 4;11	5;0 - 5;5
Cons. / *N*	12	21	17	27	7
m	mastered	mastered	mastered	mastered	mastered
n	not acquired	mastered	mastered	mastered	mastered
p	mastered	mastered	mastered	mastered	mastered
b	mastered	mastered	mastered	mastered	mastered
t	acquired	mastered	mastered	mastered	mastered
d	acquired	mastered	mastered	mastered	mastered
k	mastered	mastered	mastered	mastered	mastered
g	not acquired	mastered	acquired	mastered	mastered
f	acquired	mastered	acquired	mastered	mastered
v	not acquired	not acquired	not acquired	acquired	mastered
s	acquired	mastered	acquired	mastered	mastered
z	not acquired	not acquired	not acquired	not acquired	not acquired
ʃ	not acquired	not acquired	not acquired	not acquired	acquired
ʒ[1]	not acquired	not acquired	not acquired	not acquired	not acquired
h	not acquired	not acquired	not acquired	acquired	mastered
ɾ	not acquired	not acquired	not acquired	not acquired	not acquired
l	not acquired	mastered	mastered	mastered	mastered
j	acquired	acquired	mastered	mastered	acquired
t͡ʃ	not acquired	not acquired	acquired	acquired	acquired
d͡ʒ	not acquired	not acquired	not acquired	not acquired	not acquired

Note: **Cons.**: consonant, **shared consonants with German**, consonant mastered by age group, consonant acquired by age group, consonant not acquired by age group

[1]This consonant only occurs in loan words and according to Ege (2010) is often realised as /d͡ʒ/, even by adult speakers of Turkish

The majority of consonants not acquired or not mastered in Turkish belonged to the sound classes of fricatives and affricates. Exceptions to this were the sounds /n, g, ɾ, l/ (not acquired) and /t, d, j/ (not mastered). Additionally, mainly language-specific sounds were absent from the children's inventories whereas consonants shared across German and Turkish were already predominantly acquired by the age of 3;0 - 3;5 years (see also subsequent section).

4.1.3 Acquisition of shared consonants

In regard of the acquisition of consonants shared[10] across German and Turkish (marked in red in Table 4.1 and Table 4.2) it can be reported that many shared consonants were already acquired in both languages by the youngest age group at t1. These included: /m, p, b, t, d, k, s/ and /j/. All of them were mastered from the age of 3;6 - 3;11 years onwards. Shared consonants absent from the consonant inventories of children in the youngest age group were thereafter either acquired at the same age in both languages or earlier in German than in Turkish. The phonemes /h/ and /z/, however, were significantly later acquired in Turkish and children showed particular difficulties in the acquisition of these consonants in syllable-/word-final position – positions in which /h/ and /z/ are not permitted in German (cf. Section 1.3.1). The only shared consonants which were earlier acquired in Turkish than in German were the phonemes /g/ and /ʃ/. There was only a single syllable position in which /g/ was not acquired in German: in syllable-initial position of a weak syllable preceding a stressed syllable. In all other positions assessed (cf. Appendix S for an item list of the PLAKSS-II) it was acquired for all age groups. On the contrary, /ʃ/ was not acquired in any syllable position in German.

In general, large individual differences could be found. For only eleven children (13.1%) the acquisition of shared consonants was similar across Turkish and German. All other children had at least one shared consonant missing from their inventory in one language which they had already acquired in the other.

4.1.4 Percentages of consonants correct – adjusted

The children's PCC-A- scores were calculated from the single-word naming data following Shriberg et al.'s (1997a) criteria (see Section 3.4.3). The average PCC-A scores in German and Turkish per age group as well as the standard deviations and ranges (Min – Max) are displayed in Table 4.3.

Despite large standard deviations, statistical analyses revealed a strong correlation between the children's age and their PCC-A-scores. Older children achieved significantly higher PCC-A-scores than younger children in both languages (German: $r_s = .543$, $p < .001$; Turkish: $r_s = .426$, $p < .001$).

[10] Throughout this thesis, the term *shared consonants* refers to those consonants that are phonetically similar and share the same phonemic category across languages.

Table 4.3: PCC-A scores in German and Turkish at t1

Age group	*N*	German PCC-A				Turkish PCC-A			
		M	***SD***	**Min**	**Max**	***M***	***SD***	**Min**	**Max**
3;0 - 3;5	**12**	67.55	16.79	43.46	97.53	63.80	11.48	40.09	82.55
3;6 - 3;11	**21**	76.56	9.09	62.19	97.17	76.39	9.27	58.99	89.86
4;0 - 4;5	**17**	78.71	10.32	61.48	95.41	77.78	10.97	59.15	92.17
4;6 - 4;11	**27**	87.23**	10.69	52.30	98.94	79.90**	9.15	61.75	92.63
5;0 - 5;5	**7**	91.87*	8.23	74.20	98.94	82.01*	8.57	67.74	92.17

Note: *($p < .05$) and **($p < .01$) refer to the significance level revealed by the Wilcoxon signed-rank test when comparing the children's PCC-A-scores per age group across languages

On average, the children's PCC-A-scores were higher in their German productions than in their Turkish. Results of the Wilcoxon signed-rank test showed, though, that the differences were only significant in the two oldest age groups:

- 4;6 - 4;11: ($Mdn_{German} = 90.46$, $Mdn_{Turkish} = 82.49$), $z = -3.05$, $p = .002$, $r = -.42$[11]
- 5;0 - 5;5: ($Mdn_{German} = 95.05$, $Mdn_{Turkish} = 84.19$), $z = -2.37$, $p = .018$, $r = -.63$.

No child achieved a score of 100% correct in any language.

4.2 Acquisition of consonant clusters at t1

To address research question 1a further consonant cluster (CC) inventories were compiled per participant and analysed separately for German and Turkish. The scoring criteria applied for these analyses are outlined in Section 3.4.2.

4.2.1 Acquisition of German onset consonant clusters

Table 4.4 displays the acquisition (75%-criterion) and mastery (90%-criterion) of German onset CCs comparing the results of the three scoring criteria. According to the first scoring criterion the first two German onset CCs to be acquired were /gl/ and /kl/ at the age band of 4;0 - 4;5 years. The number of acquired onset CCs significantly increased with age ($r_s = .511$, $p < .001$) so that by the age of 5;0 - 5;5 years, children had acquired ten out of 23 onset CCs and had mastered five of them.

[11] Throughout this thesis, effect sizes were interpreted as follows: $r = .10$ (small), $r = .30$ (medium), $r = .50$ (large) in line with Cohen (1988).

Table 4.4: Acquisition and mastery of German onset CCs – comparison of three different scoring criteria at t1

Age group		3;0-3;5	3;6-3;11	4;0-4;5	4;6-4;11	5;0-5;5
CC	Cr. \ N	12	21	17	27	7
fl	1	yellow	yellow	yellow	dark blue	dark blue
	2	yellow	yellow	yellow	dark blue	dark blue
	3	yellow	yellow	light blue	dark blue	dark blue
fʁ	1	yellow	yellow	yellow	yellow	light blue
	2	yellow	yellow	yellow	yellow	light blue
	3	yellow	yellow	yellow	light blue	dark blue
bl	1	yellow	yellow	yellow	light blue	dark blue
	2	light blue	light blue	yellow	light blue	dark blue
	3	yellow	yellow	light blue	light blue	dark blue
bʁ	1	yellow	yellow	yellow	yellow	light blue
	2	yellow	yellow	yellow	yellow	light blue
	3	yellow	yellow	yellow	light blue	dark blue
dʁ	1	yellow	yellow	yellow	yellow	light blue
	2	yellow	yellow	yellow	yellow	light blue
	3	yellow	yellow	yellow	light blue	dark blue
gl	1	yellow	yellow	light blue	light blue	dark blue
	2	yellow	yellow	light blue	light blue	dark blue
	3	yellow	yellow	dark blue	light blue	dark blue
gʁ	1	yellow	yellow	yellow	yellow	light blue
	2	yellow	yellow	yellow	yellow	light blue
	3	yellow	yellow	yellow	light blue	dark blue
tʁ	1	yellow	yellow	yellow	yellow	light blue
	2	yellow	yellow	yellow	yellow	light blue
	3	yellow	yellow	light blue	yellow	dark blue
kl	1	yellow	yellow	light blue	dark blue	dark blue
	2	yellow	light blue	light blue	dark blue	dark blue
	3	yellow	light blue	dark blue	dark blue	dark blue
kʁ[1]	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	yellow
	3	yellow	yellow	yellow	yellow	yellow
kn	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	light blue
	3	yellow	yellow	yellow	yellow	light blue
kv	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	yellow
	3	yellow	yellow	yellow	yellow	yellow
ʃl	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	yellow
	3	yellow	light blue	light blue	dark blue	dark blue
ʃʁ	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	yellow
	3	yellow	yellow	light blue	dark blue	light blue
ʃm	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	yellow
	3	yellow	yellow	light blue	dark blue	dark blue
ʃn	1	yellow	yellow	yellow	yellow	yellow
	2	yellow	yellow	yellow	yellow	yellow

Age group		3;0-3;5	3;6-3;11	4;0-4;5	4;6-4;11	5;0-5;5
CC	Cr. \ N	12	21	17	27	7
	3	not acquired	acquired	mastered	mastered	mastered
ʃv	1	not acquired	not acquired	not acquired	not acquired	not acquired
	2	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	acquired	mastered	acquired	mastered
ʃp	1	not acquired	not acquired	not acquired	not acquired	not acquired
	2	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	acquired	acquired	mastered
ʃt	1	not acquired	not acquired	not acquired	not acquired	not acquired
	2	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	acquired	acquired	acquired	mastered
ʃpʁ	1	not acquired	not acquired	not acquired	not acquired	not acquired
	2	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	not acquired	not acquired	acquired
ʃtʁ	1	not acquired	not acquired	not acquired	not acquired	not acquired
	2	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	not acquired	not acquired	acquired
pfl[2]	1	not acquired	not acquired	not acquired	acquired	acquired
	2	not acquired	not acquired	not acquired	acquired	mastered
	3	not acquired	acquired	not acquired	mastered	mastered
tsv[3]	1	not acquired	not acquired	not acquired	not acquired	not acquired
	2	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	acquired	acquired	mastered	mastered

Note: **Cr.**: criterion, **criterion 1**: CC considered to be correct if it is produced phonemically correct, **criterion 2**: CC is considered to be correct if the cluster elements are produced correctly but split up by an epenthetic vowel/Schwa, **criterion 3**: CC is considered to be correct if the number of cluster elements is produced correctly, CC mastered by age group, CC acquired by age group, CC not acquired by age group

[1]/kʁ/ is the only CC that is assessed in a weak syllable (i.e., /kʁokoˈdiːl/). This may have affected the accuracy of production for this CC and its absence from the children's inventories across all age groups is likely to be an artefact of the test item rather than an acquisition issue.

[2]In word-initial position /p͡f/ is usually deaffricated to /f/ in Northern German, thus this realisation was also accepted as phonemically correct (see Sections 1.3.1.4 and 3.4.1).

[3]The presented age of acquisition for this CC is potentially not representative for this sample since 87% of the children could not produce the item spontaneously and had to be prompted by delayed ($n = 37$) or direct imitation ($n = 36$). In addition, /t͡sv/ generally only occurs in a very limited number of child appropriate words.

All onset CCs acquired by an age group, except for /gl/ in the 4;0 - 4;5-year-olds, contained sounds the children had already acquired as singletons at this age (cf. Section 4.1.1). Onset clusters containing the liquid /l/, except for /ʃl/, were acquired before those including the liquid /ʁ/. All other onset CC, the majority being /ʃ/-CC, were still absent from the children's inventories at the age of 5;0 - 5;5 years.

If children had not acquired a German onset CC phonemically correct they either reduced the CC to one (or two) elements, split it up by an epenthetic vowel/schwa or simplified the production by substituting one or both elements by another sound but often keeping the correct syllable shape. In only three instances the whole onset CC was deleted.

When accepting the cross-linguistic transfer of inserting a vowel/schwa into an onset CC as a correct production (i.e., scoring criterion 2), the children had acquired their first onset

CC at the age of 3;0 - 3;5 years (i.e., /bl/) and had acquired eleven German onset CCs (five of them also mastered) by the age of 5;0 - 5;5 years.

The third scoring criterion, with which the acquisition of the underlying representation of the syllable structure of a CC was evaluated, revealed that children could produce two-element German onset CCs already at the age of 3;6 - 3;11 years. Further, children in the oldest age group (i.e., 5;0 - 5;5 years) had acquired the syllable structures of all assessed two- and three-element onset clusters but /kv/ and /kʁ/.

4.2.2 Acquisition of German coda consonant clusters

According to the first scoring criterion Turkish-German bilingual children had already acquired three German coda CCs at the age of 3;0 - 3;5 years (see Table 4.5). Two of these were also mastered by this age group. Generally, the acquisition of coda CCs followed a developmental pattern with older children having acquired more coda CCs than younger children, $r_s = .482$, $p < .001$.

Table 4.5: Acquisition and mastery of German coda CCs – comparison of two different scoring criteria at t1

Age group		3;0 - 3;5	3;6 - 3;11	4;0 - 4;5	4;6 - 4;11	5;0 - 5;5
CC	**Cr.** / ***N***	**12**	**21**	**17**	**27**	**7**
lç	1					
	3					
lp	1					
	3					
lt	1					
	3					
ns	1					
	3					
nt	1					
	3					
ŋk	1					
	3					
st	1					
	3					
çt	1					
	3					
nst	1					
	3					

Note: **Cr.**: criterion, **criterion 1**: CC is considered to be correct if it is produced phonemically correct, **criterion 3**: CC is considered to be correct if the number of cluster elements is produced correctly, CC mastered by age group, CC acquired by age group, CC not acquired by age group

Coda CCs were only then acquired phonemically correct if their elements were acquired as singletons. The clusters /ŋk/ and /çt/, however, form exceptions (cf. Section 4.1.1).

Coda CCs formed by nasal/liquid + stop were acquired before those consisting of fricative + stop or liquid + fricative. By the age of 5;0 - 5;5 years three coda CCs were still absent from the children's inventories (i.e., /lç/, /st/, /nst/). If children had not acquired a German coda CC phonemically correct they either reduced the CC to one (or two) elements or simplified the production by substituting one or both elements by another sound but often keeping the correct syllable shape (i.e., scoring criterion 3). Apart from an earlier acquisition of the syllable structure of /lp/ and an earlier mastery of /ns/ and /çt/, the results of scoring criterion 3 present a similar acquisition pattern as those for criterion 1. There were only three instances of vowel epentheses in coda CCs, hence, data with criterion 2 is not presented in Table 4.5. Deletions of coda CCs did not occur.

4.2.3 Acquisition of Turkish coda consonant clusters

Only two of the assessed Turkish CCs were acquired phonemically correct (i.e., scoring criterion 1) by the children, with /nk/ from the youngest age group and /lp/ from the age group of 3;6 - 3;11 years onwards (see

Table 4.6). Both clusters contained phonemes which were already acquired as singletons by these age groups. Despite only two clusters being acquired, a developmental acquisition trend was found with more clusters being acquired by the older children, $r_s = -.296$, $p = .006$.

Table 4.6: Acquisition and mastery of Turkish coda CCs – comparison of two different scoring criteria at t1

Age group		3;0 - 3;5	3;6 - 3;11	4;0 - 4;5	4;6 - 4;11	5;0 - 5;5
CC	**Cr.** / ***N***	**12**	**21**	**17**	**27**	**7**
lp	1	not acquired	acquired	acquired	acquired	mastered
	3	not acquired	acquired	acquired	mastered	mastered
ɾt	1	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	not acquired	not acquired	not acquired
ɾk	1	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	not acquired	not acquired	not acquired
nt͜ʃ	1	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	acquired	acquired	mastered	acquired
nk	1	acquired	acquired	acquired	mastered	mastered
	3	acquired	mastered	mastered	mastered	mastered
ft	1	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	not acquired	not acquired	not acquired

Note: **Cr.**: criterion, **criterion 1**: CC considered to be correct if it is produced phonemically correct, **criterion 3**: CC is considered to be correct if the number of cluster elements is produced correctly, CC mastered by age group, CC acquired by age group, CC not acquired by age group

Turkish coda CCs that were not acquired by an age group were either affected by reductions or substitutions of one or two CC members. Vowel epentheses only occurred once in the whole data set, thus scoring criterion 2 was not included in Table 4.6. Deletions of whole coda CCs only occurred in three instances. Results of criterion 3 revealed that children had acquired the underlying syllable shape of three coda CCs by the age of 3;6 - 3;11 years and mastered by the age of 4;6 - 4;11 years (i.e., /lp/, /nt͜ʃ/, /ft/).

4.3 Phonological variations from adult-like speech at t1

To qualitatively describe and understand the children's developmental realisations of a certain consonant or syllable structure and thus to address research question 1b, children's phonological variations from adult-like speech were evaluated.

The following section starts with an overview of the children's overall number of PhonVar from adult-like speech. Then, the number of InfVar among the PhonVar produced is reported, before the (number of different) types of phonological patterns and the ages of their suppression are outlined. Finally, those variations which were considered to be due to a CLT are presented for both languages.

4.3.1 Number of phonological variations

To report the children's overall degree of phonological variability their number of PhonVar from the adult pronunciation were calculated for both languages (see Table 4.7). For this, variations resulting from a CLT were not included.

There was a significant negative moderate correlation between the children's age and the number of PhonVar produced in German, $r_s = -.575$, $p < .001$ and Turkish, $r_s = -.418$, $p < .001$. On average, children in the youngest age group (i.e., 3;0 - 3;5 years) produced the largest number of PhonVar across languages which gradually decreased in the older age groups. Standard deviations were very high, especially in the youngest age group, indicating a large variability in the children's performances at this age with a range from 8 to 159 PhonVar in German and 42 to 167 in Turkish. No child had completely achieved adult-like speech (i.e., produced zero PhonVar).

Table 4.7: Number of PhonVar produced in German and Turkish per age group at t1

Age group	*N*	PhonVar German				PhonVar Turkish			
		M	*SD*	Min	Max	*M*	*SD*	Min	Max
3;0 - 3;5	**12**	98.08	50.76	8	159	87.75	32.90	42	167
3;6 - 3;11	**21**	70.95	27.95	10	121	56.14	22.65	25	97
4;0 - 4;5	**17**	63.06	31.72	12	126	54.88	30.53	19	104
4;6 - 4;11	**27**	35.93	32.69	8	141	48.15	24.31	17	96
5;0 - 5;5	**7**	21.86	19.59	4	55	43.29	24.21	18	81

Up to the age of 4;0 - 4;5 years, the mean number of PhonVar was higher in the children's German productions than in their Turkish (e.g., 98.08 vs. 87.75 in the youngest age group). The reverse was the case for the age groups 4;6 - 4;11 and 5;0 - 5;5 years. The overall difference between the number of PhonVar in German ($Mdn = 54$) and Turkish ($Mdn = 51$) was not statistically significant, Wilcoxon signed-rank test: $z = -.26$, $p = .798$, $r = -.02$. However, the differences in the two languages in the age groups 3;6 - 3;11, 4;6 - 4;11 and 5;0 - 5;5 were all significant at the $p < .05$ level with small to large effect sizes:

- 3;6 - 3;11: ($Mdn_{German} = 72$, $Mdn_{Turkish} = 49$), $z = -2.41$, $p = .014$, $r = -.37$
- 4;6 - 4;11: ($Mdn_{German} = 28$, $Mdn_{Turkish} = 46$), $z = -2.03$, $p = .041$, $r = -.28$
- 5;0 - 5;5: ($Mdn_{German} = 16$, $Mdn_{Turkish} = 35$), $z = -2.37$, $p = .016$, $r = -.63$.

For all these analyses, it should be borne in mind that the number of items in the TPA ($N = 70$) is smaller than in the PLAKSS-II ($N = 96$).

Additionally, a significant correlation was found between the children's PCC-A scores and the number of PhonVar (only including those affecting consonants and consonant clusters) in both languages. The larger the number of PhonVar was, the lower the PCC-A-score, German: $r_s = -.979$, $p < .001$ and Turkish: $r_s = -.687$, $p < .001$.

4.3.2 Number of infrequent variants

In order to analyse the children's degree of speech perfection (i.e., their developmental level in the acquisition of adult-like speech; Fox, 2016), the number of InfVar among the PhonVar was calculated. As for the number of PhonVar the number of InfVar excludes those resulting from CLT. Infrequent variants consisted of (a) item-related specific variations – i.e., variations that only occurred once in many children and only with this particular item (e.g., final consonant deletion in *Trampolin* (trampoline) [ˈtʁampoli] or metathesis in *Zebra* (zebra) [ˈtsʁeːba]) and (b) those that were simply infrequent versions of typically known phonological patterns.

Infrequent variants constituted 51.3% (German) and 56.8% (Turkish) of the PhonVar in the youngest age group (i.e., 3;0 - 3;5 years) and the proportion continuously increased to 84.1% (German) and 78.5% (Turkish) in the oldest age group (5;0 - 5;5 years). The average numbers of InfVar in German and Turkish as well as the respective standard deviations and ranges (Min and Max) per age group are displayed in Table 4.8.

Table 4.8: Number of InfVar in German and Turkish produced per age group at t1

Age group	***N***	**InfVar German**				**InfVar Turkish**			
		M	***SD***	**Min**	**Max**	***M***	***SD***	**Min**	**Max**
3;0 - 3;5	**12**	40.00	18.54	8	68	45.83	11.49	23	60
3;6 - 3;11	**21**	38.86	13.39	10	62	35.67	10.63	19	59
4;0 - 4;5	**17**	30.82	9.43	12	47	32.59	13.58	16	58
4;6 - 4;11	**27**	22.70	14.75	7	67	31.70	13.56	11	62
5;0 - 5;5	**7**	14.43	9.78	4	28	31.14	13.93	15	56

Children's production of InfVar in German and Turkish was moderately correlated with their age, $r_s = -.575$, $p < .001$ and $r_s = -.418$, $p = .001$ respectively, with younger children producing significantly more InfVar than older children.

The average number of InfVar was similar in both languages for most age groups. Children in the age group of 4;6 - 4;11 years, however, produced significantly fewer InfVar in German ($Mdn = 19$) than in Turkish ($Mdn = 28$), Wilcoxon signed-rank test: $z = -2.36$, $p = .018$, $r = -.32$. The same applies to the difference in the oldest age group, Wilcoxon signed-rank test: ($Mdn_{German} = 15$, $Mdn_{Turkish} = 28$), $z = -2.37$, $p = .018$, $r = -.63$. Still, there was a high variability in the children's performances (see *SDs* in Table 4.8).

4.3.3 Quantity and quality of phonological patterns

This section reports the average number of phonological pattern types in German and Turkish per age group as well as the individual patterns observed and their ages of suppression in both languages.

4.3.3.1 Number of different pattern types

The proportion of phonological patterns among the children's variations from adult-like speech was 50.6% (German) and 45.1% (Turkish) in the 3;0 - 3;5-year-olds and gradually decreased to 19.9% (German) and 33.2% (Turkish) in the 5;0 - 5;5-year-olds.

Table 4.9 demonstrates the children's average number of different phonological pattern types per age group in German and Turkish.

Table 4.9: Average number of phonological pattern types in German and Turkish observed at t1

Age Group	*N*	Pattern types German				Pattern types Turkish			
		M	*SD*	Min	Max	*M*	*SD*	Min	Max
3;0 - 3;5	**12**	6.08	3.40	0	11	4.67	2.74	2	10
3;6 - 3;11	**21**	3.62**	2.01	0	8	2.33**	1.56	0	5
4;0 - 4;5	**17**	2.82	2.60	0	10	2.35	2.09	0	7
4;6 - 4;11	**27**	1.19*	1.90	0	8	1.96*	1.53	0	5
5;0 - 5;5	**7**	0.71	1.25	0	3	1.29	1.38	0	3

Note: *Difference in the number of phonological pattern types across languages is significant at the $p < .05$ level, **Difference in the number of phonological pattern types across languages is significant at the $p < .01$ level

Younger children exhibited a significantly larger number of pattern types than older children (German: $r_s = -.595$, $p < .001$, Turkish: $r_s = -.361$, $p < .001$). However, large standard deviations and ranges emphasise a high variability in the children's use of phonological pattern types. Some children in the youngest age group, for example, did not produce any phonological pattern type in German whereas others produced 13. A comparable range was found for Turkish.

Up to the age group of 4;0 - 4;5, the children produced more pattern types in German than in Turkish on average. For the remaining two age groups the reverse was the case. However, the differences between the number of pattern types in German and Turkish were only statistically significant for the older three-year-olds and the older four-year-olds, Wilcoxon signed-rank tests:

- 3;6 - 3;11: ($Mdn_{German} = 6$, $Mdn_{Turkish} = 3$), $z = -2.84$, $p = .003$, $r = -.62$
- 4;6 - 4;11: ($Mdn_{German} = 9.5$, $Mdn_{Turkish} = 5$), $z = -2.07$, $p = .037$, $r = -.40$.

4.3.3.2 Phonological pattern types and their age of overcome

All phonological patterns that could be observed in at least 15% of an age group are presented in the following. For comparison purposes with previous research, additionally those patterns which occurred in only 10 - 14% of the children in an age group are reported (cf. Section 3.4.4). The presence of the observed phonological patterns in the children's speech needs to be interpreted with great caution and does not necessarily imply their "definite" presence in typical development. Especially, patterns present in particularly small age groups (i.e., the youngest and oldest age group) may yield to misleading interpretations.

Following Grunwell (1985) the observed phonological pattern types were divided into structural and systemic simplifications. Structural simplifications pertain to phonological changes in the syllable or word structure, whereas systemic simplifications apply to substitutions of single sounds which do not alter the overall structure of a syllable or word. For examples of all phonological patterns and types of CLT referred to in this research see Appendix U.

German

Structural simplifications were suppressed at the latest by 5;0 years (see Table 4.10). The most frequent pattern affecting the syllable structure was the reduction of initial CCs followed by the deletion of weak syllables. Unlike reductions of initial CCs, weak syllable deletion never occurred in every possible item (range of affected items: 5 - 10). The remaining two structural patterns (i.e., final CC reduction and intrusion of a consonant) only occurred in few children and never affected all possible items in the test.

Table 4.10: Phonological patterns produced in German at t1

Pattern / Age group	3;0 - 3;5 *n* (%)	3;6 - 3;11 *n* (%)	4;0 - 4;5 *n* (%)	4;6 - 4;11 *n* (%)	5;0 - 5;5 *n* (%)
Structural simplifications					
Initial CC reduction[1]	10 (83)	17 (81)	10 (59)	8 (30)	
Weak syllable deletion	5 (42)	8 (38)	3 (18)		
Final CC reduction[2]	4 (33)	4 (19)	2 (12)	3 (11)	
Intrusive consonant	4 (33)	3 (14)	2 (12)		
Systemic simplifications					
Fronting /ʃ/	9 (75)	13 (62)	7 (41)	7 (26)	2 (29)
Deaffrication /t͜s, p͜f/	3 (25)	9 (43)	6 (35)		1 (14)
Devoicing[3]	5 (42)	2 (10)	3 (18)		1 (14)
Stopping Fric. & Affr.	6 (50)	5 (24)			
Assimilation	5 (42)	3 (14)			
Fronting /k, g, ŋ/	3 (25)		2 (12)	3 (11)	1 (14)
Backing /t, d, n/	3 (25)				
Metathesis	2 (17)	2 (10)			
Voicing[3]	3 (25)				
Labialisation	2 (17)				
Fronting /ç/		3 (14)	2 (12)		
Allophonic use of sibilants			2 (12)		

Note: **[1]**includes syllable-initial consonant clusters, **[2]**includes syllable-final consonant clusters, **[3]**(de-)voicing also occurs in a range of regional varieties of German (Fox-Boyer, 2014c; Fox, 2007; Wiese, 2011) although not in the ones where the data were collected, **Fric.**: fricatives, **Affr.**: affricates, produced by 15% or more of an age group, blank cells: pattern occurred in 0 - 9.9% of an age group

Except for fronting of velars and fronting of /ʃ/ all systemic simplifications were overcome at the latest by 4;6 years. Whereas fronting of /ʃ/ was shown by 45% of the whole sample at t1 and usually affected many targets (in 10% of the cases all 20 target items), fronting of velars was relatively rare, i.e., occurred in less than 11% of the sample and usually only affected a few targets.

Deaffrication was the second most frequent systemic simplification, affecting /t͡s/ slightly more frequently than /p͡f/. If it was present in a child it was usually evident in all possible items. Devoicing often affected word-initial /g/ and /z/ and only in two cases occurred more than ten times per participant. In contrast, fronting of /ç/ usually affected all items containing this phoneme.

Stopping was usually found for fricatives only. In general, this pattern occurred very inconsistently and only in a very small subset of children per age group. Assimilation was observable for a small number of items per child (i.e., no more than eleven), thus was also an inconsistent pattern. The pattern could most frequently be found in the word *Eichhörnchen* (squirrel) /ˈaɪçhøɐnçən/ which was produced as [ˈaɪnhøɐnçən].

The patterns backing of /t, d, n/, metathesis, voicing, labialisation and allophonic use of sibilants were not very frequent and were never produced consistently. A child who backed /t, d, n/, for example, did not show this pattern in every phonetically possible occasion in the test. An exception to this was the allophonic use of sibilants as it occurred to a maximum of 43 tokens per child out of 50 possible occurrences.

<u>Turkish</u>

Structural simplifications in Turkish were generally overcome at an early age, except for the deletion of syllable-final consonants (see Table 4.11). This pattern was the most frequent structural simplification and mainly affected the sounds /n, l, ɾ, j, h/. Final consonant deletions were present in two age groups, the 3;0 - 3;5-year-olds and the 4;6 - 4;11-year-olds. In both age groups word-final /ɾ/ was deleted the most. Intrusions of consonants as well as initial consonant deletions were very rare patterns.

In contrast to structural simplifications, systemic simplifications were mainly overcome at a later age, with fronting of /ʃ, ʒ, t͡ʃ, d͡ʒ/, devoicing and liquid deviation being the most prominent phonological patterns. Whereas fronting of postalveolar fricatives/affricates often occurred systematically (i.e., affecting all or most of the items containing these sounds), liquid deviation predominantly affected /ɾ/ in syllable-initial (37%) and word-final positions (33.6%). Devoicing was only produced in a subset of items and occurred to a maximum of 20 occurrences per participant.

Table 4.11: Phonological patterns produced in Turkish at t1

Pattern \ Age group	3;0 - 3;5 *n* (%)	3;6 - 3;11 *n* (%)	4;0 - 4;5 *n* (%)	4;6 - 4;11 *n* (%)	5;0 - 5;5 *n* (%)
Structural simplifications					
SF consonant deletion	8 (67)	11 (52)	7 (41)	9 (33)	2 (29)
Final cons. deletion	2 (17)			5 (19)	
Final CC reduction[1]	2 (17)				
Intrusive consonant	2 (17)				
Initial consonant deletion	2 (17)				
Systemic simplifications					
Fronting /ʃ, ʒ, t͡ʃ, d͡ʒ/	8 (67)	13 (62)	7 (41)	9 (33)	2 (29)
Devoicing[2]	6 (50)	8 (38)	5 (29)	9 (33)	3 (43)
Liquid deviation /r/ → [l]	5 (42)	4 (19)	4 (24)	6 (22)	2 (29)
Assimilation	8 (67)	6 (29)	4 (24)	6 (22)	1 (14)
Deaffrication /t͡ʃ, d͡ʒ/	2 (17)	4 (19)	3 (18)	5 (19)	
Gliding /r/ → [j, w]	3 (25)		4 (24)	3 (11)	
Stopping Fric. & Affr.	5 (42)		3 (18)		
Metathesis	2 (17)	2 (10)		3 (11)	
Backing /s, z/	2 (17)				
Word-final devoicing					1 (14)

Note: **SF**: syllable-final, **[1]**includes syllable-final consonant clusters, **[2]**excluding devoicing of word-intial alveolar and velar stops as these could be dialectal/accentual vairations (Brendemoen, 1998), **Fric.**: fricatives, **Affr.**: affricates, produced by 15% or more of an age group, blank cell: pattern occurred in 0 - 9.9% of an age group

Two further frequent systemic simplifications were assimilation and deaffrication. Although many children across all age groups showed assimilations in their speech, this pattern only occurred inconsistently, with a maximum occurrence of 14 times per participant. In most cases children exhibited progressive assimilations, i.e., a sound was assimilated in place, voice or manner to a later occurring sound in the word. Deaffrication was occasionally present in all items containing affricates; however, the majority of children only produced five to six instances of this pattern in the test.

Gliding of /r/ predominantly affected onset positions (71.7%) with the substitutional consonant mainly being /j/. It was rarely used as a consistent pattern and usually co-existed with liquid deviation, i.e., depending on the word position children substituted the /r/ by /j, w/ or /l/.

Metathesis was only present in a maximum of eight words per child and predominantly affected tri-syllabic words. Stopping could more often be observed for fricatives than for affricates and it usually occurred in a small number of instances per child, i.e., five to seven. The patterns backing of /s, z/ and devoicing of consonants in word-final positions were very rare and were only produced to a maximum of eight and five instances respectively.

Comparison of German and Turkish

Although a large number of phonological patterns could have occurred in both languages, the children did not necessarily produce them in both. Fronting of velars, backing of /t, d, n/, weak syllable deletion and labialisation, for example, only occurred in German, whereas backing of alveolar sibilants, deletion of (syllable-) final and initial consonants were only present in Turkish.

In addition, patterns observed in both languages (i.e., final CC reduction, fronting of postalveolar fricatives, assimilation, devoicing, stopping, deaffrication, metathesis) were partly overcome at different ages. Assimilation, for example, was only present until the age of 3;11 years in German but occurred until the age of 4;11 years in Turkish. Thus, the phonological patterns the children exhibited in German at t1 were both similar and dissimilar from the ones produced in Turkish.

4.3.4 Cross-linguistic transfer

This section reports on the type of patterns attributable to a cross-linguistic transfer. These included phonological as well as a few phonetic variations (marked with an asterisk in subsequent tables) that were subject to an interaction of the two languages. Although some types of CLT patterns occurred in less than ten percent of an age group all observed patterns are reported to demonstrate the variety of interactions possible in the language combination of Turkish and German.

German

Cross-linguistic transfer mainly affected the production of the consonant /ʁ/ which was either substituted by the Turkish equivalent (i.e., /r/) or affected by its common simplification patterns (i.e., liquid deviation (/ʁ/ → [l]) and gliding (/ʁ/ → [j]; see Table 4.12). Both substitutions are unusual in German. Substitutions of the /ʁ/, except for gliding, mostly occurred consistently (i.e., were produced in every possible phonetic occasion in the test). However, children did not show them beyond the age of 4;11 years.

Table 4.12: Cross-linguistic transfer observed in German at t1

Pattern \ Age group	3;0 - 3;5 *n* (%)	3;6 - 3;11 *n* (%)	4;0 - 4;5 *n* (%)	4;6 - 4;11 *n* (%)	5;0 - 5;5 *n* (%)
Structural simplifications					
Vowel epenthesis in CCs	3 (25)	6 (29)	2 (12)	1 (4)	
BPR of /ŋ/ → [ŋg]*			1 (4)		
Systemic simplifications					
/ʁ/ → [l]	2 (17)	4 (19)	1 (6)	3 (11)	1 (14)
/ʁ/ → [ɾ]*	1 (8)	4 (19)	4 (24)	3 (11)	
/ç/ → [ʃ]*[1]				1 (4)	1 (14)
/ʁ/ → [r]*				1 (4)	
/ʁ/ → [j]				1 (4)	

Note: **BPR**: Biphonematic realisation, *****phonetic patterns, produced by 15% or more of an age group, produced by less than 10% of the age group, blank cells: pattern did not occur in this age group

[1]This pattern also occurs as a dialectal variation in the Rhineland region where some of the participants of this study originated from. However, it has also been frequently reported to be a CLT phenomenon in Turkish-German bilingual speakers (Androutsopoulos, 2001; Bücker, 2007).

Further patterns ascribable to a transfer from Turkish to German concerned sounds or syllable structures that do not occur in Turkish (i.e., CCs, /ç/ and /ŋ/). All were simplified by structures or phonemes included in the Turkish phonological system which resulted in a resembling pronunciation (e.g., epenthesis of a vowel in CCs; see Sections 4.2.1 and 4.2.2 for more details).

Turkish

Cross-linguistic transfer was generally less frequent in Turkish than in German (i.e., not occurring in at least 15% of an age group) and included variations of the tap and the regular stress pattern (see Table 4.13). The Turkish /ɾ/ was occasionally replaced by the German equivalent or was affected by German phonotactic restrictions (i.e., a vocalisation of /ɾ/ → [ɐ] in (syllable-) final position and as C_1 in word-final consonant clusters). In case of an incorrect use of lexical stress a trochaic stress pattern was applied, which is the most prominent one in German (e.g., /gʌzɛˈtɛ/ → [gʌˈzɛtɛ]).

Table 4.13: Cross-linguistic transfer observed in Turkish at t1

Pattern \ Age group	3;0 - 3;5 *n* (%)	3;6 - 3;11 *n* (%)	4;0 - 4;5 *n* (%)	4;6 - 4;11 *n* (%)	5;0 - 5;5 *n* (%)
Structural simplifications					
Incorrect stress pattern		3 (14)		3 (11)	
Systemic simplifications					
/ɾ/ → [ʁ]*		1 (5)		2 (7)	1 (14)
Vocalisation of /ɾ/		1 (5)			

Note: produced by less than 10% of the age group, blank cells: pattern did not occur in this age group, *phonetic pattern

Overall, only one out of the 39 children who produced CLT demonstrated evidence of a bi-directional CLT. All other children either only demonstrated CLT from German to Turkish or vice versa. However, many of these children showed CLT instances in both languages but the frequency of occurrence of these instances was below the cut-off (i.e., fewer than five times) so that they are not listed here.

4.4 Summary – Phonological acquisition at t1

The cross-sectional data on the children's phonological acquisition in Turkish and German revealed that Turkish-German bilingual children had already acquired many consonants in both of their languages at the age of 3;0 - 3;5 years. Further, a developmental trend with older children having acquired more consonants than younger children was observed in both languages. However, the children's consonant inventories were still incomplete at the age of 5;5 years in both Turkish and German. Late acquired or absent consonants mainly belonged to the sound classes of fricatives and affricates and/or comprised language-specific sounds. Shared consonants were mostly acquired at the same age in both languages or slightly earlier in German than in Turkish. Despite a large inter-participant variability, no child achieved a percentage of consonants correct score of 100% in either language at t1, though, some children in every age group approached this score by achieving scores above 90%. On average, consonant accuracy was higher in German than in Turkish.

Consonant cluster inventories in both languages demonstrated that Turkish-German bilingual children did not have any German onset CCs acquired by the age of 3;0 - 3;5 years but a few coda clusters in German and one in Turkish. Similar to the children's acquisition of single consonants, their onset and coda CC inventories remained incomplete at the age of 5;5 years for both languages. However, the data also revealed that children had acquired the underlying syllable structure of the CCs a lot earlier than their phonemically correct realisation. Clusters with which the children struggled the most mainly included phonemes that had also not been acquired as singletons (e.g., /ʃ/).

When consonants or CCs had not been acquired, the children demonstrated a range of PhonVar from adult-like speech to substitute or omit them. Those included InfVar, structural and systemic phonological patterns as well as phonological CLT. Most frequently CCs and late acquired consonants such as /ʃ, ɾ/ were affected by these. There were both similarities and differences in the types of PhonVar observed in the children's German and Turkish productions. These were found to decrease with age and on average to occur more frequently in Turkish than in German.

4.5 Factors influencing children's phonological performance

In the previous three sections (i.e., 4.1, 4.2 and 4.3) of this chapter the children's phonological performances were described. To find out whether these were influenced by other variables than chronological age (which was found to correlate with children's phonological performances across all measures and in both languages) and to answer research question 2, differences in the children's phonological performances regarding gender, proportion of language input and output as well as language proficiency were computed. The AoA of each language was not included as a variable in this analysis for reasons outlined in Section 3.2.3.2).

It was decided to explore differences for two reasons: (a) since categorical variables were used for most of the potential influential factors (i.e., gender, proportion of input, and proportion of output) and (b) because correlations between phonological skills and input/output patterns as well as language experience were investigated in previous studies, so that differences offered the possibility to go one step beyond exploring relationships.

As outcome variables, the three holistic measures *number of PhonVar, number of InfVar* and *PCC-A-scores* for both languages were used. Due to their holistic nature these variables were preferred over more specific ones (e.g., consonant inventory) as they provide insight into the overall degree of variability in children's phonological productions.

Due to the unequal sample size across age groups non-parametric tests were used for all analyses. Since these tests do not allow for controlling of additional variables, each of the following sections begins with a short paragraph reporting on any age differences present between the individual groups/categories formed per influential factor.

4.5.1 Effects of gender on phonological performance

To explore any gender-related effects on children's phonological performances the boys' and girls' number of PhonVar, number of InfVar and PCC-A-scores were analysed and compared in both languages.

<u>Age differences in the gender variable</u>

Mann-Whitney tests revealed no significant age differences between the two comparing groups, i.e., boys (*Mdn* = 53.00 months) and girls (*Mdn* = 49.00 months), $U = 783.50$, $z = -.78$, $p = .441$, $r = -.09$.

Gender-related differences

The average performances of boys and girls on the different phonological measures suggested weaker phonological skills in boys across all three outcome measures and both languages (see Table 4.14). Mann-Whitney tests, however, did not reveal any significant gender differences except for the number of InfVar in Turkish which was higher for boys. Since the effect was small ($r = -.22$), gender did not seem to have a generally significant influence on the children's phonological performances.

Table 4.14: Performance on phonological outcome measures at t1 by gender

Outcome measure	Male (*n* = 37)			Female (*n* = 47)			*Mann-Whitney**			
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*U*	*z*	*p*	*r*
PhonVar German	61.11	43.77	45.00	55.34	37.89	56.00	817.50	-.47	.643	-.05
PhonVar Turkish	62.51	27.04	64.00	52.23	30.34	42.00	659.00	-1.90	.058	-.21
InfVar German	31.81	17.12	30.00	28.87	14.95	31.00	814.00	-.50	.620	-.05
InfVar Turkish	38.03	14.00	40.00	32.34	12.15	28.00	641.50	-2.06	.040	-.22
PCC-A German**	79.48	14.36	84.45	81.15	12.80	80.92	815.50	-.49	.630	-.05
PCC-A Turkish**	74.18	10.78	72.35	78.27	11.57	82.49	665.50	-1.84	.066	-.20

Note: *****: refers to differences between males and females, **: values presented correspond to %, significant difference

4.5.2 Effects of the proportion of language input on phonological performance

To investigate the influence of the proportional distribution of language input on children's phonological performances between-group and within-group differences were examined. Between-group differences concern comparisons of the children's phonological performances in both languages between the different categories of language input (i.e., MoreG, Equal and MoreT; see Section 3.2.3). Within-group differences refer to comparisons of children's performances across languages within each of the different language input categories.

4.5.2.1 Differences between language input categories regarding phonological outcome measures

As for the effects of gender, reports on any age differences between the language input categories precede those for input-related effects on phonological performances.

Age differences in the language input variable

A Kruskal-Wallis test revealed that children in the three language input categories differed significantly on age, $H(2) = 8.465$, $p = .015$. A follow-up with Mann-Whitney tests using a Bonferroni correction (cut-off for significance $p < .017$) showed that children with more language input in German (*Mdn* = 56.50) were significantly older than children who received an equal amount of language input in both languages (*Mdn* = 48.00) and those who had more input in Turkish (*Mdn* = 47.00), $U_{MoreG-Equal} = 219.00$, $z = -2.48$, $p = .013$, $r = -.33$, $U_{MoreG-MoreT} = 94.00$, $z = -2.64$, $p = .007$, $r = -.42$). Children in the categories Equal and MoreT, however, did not differ significantly on age, $U = 277.00$, $z = -.06$, $p = .955$, $r = -.01$.

Language input-related differences

Across all three outcome measures, more input in German resulted in on average better phonological skills in German than an equal amount of input in both languages and more input in Turkish (see Table 4.15).

Table 4.15: Phonological outcome measures at t1 by proportion of language input

Outcome measure	MoreG (*n* = 22)			Equal (*n* = 33)			MoreT (*n* = 17)		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
PhonVar German	31.77	24.88	27.00	58.91	43.75	55.00	75.45	31.60	78.00
PhonVar Turkish	48.32	24.47	46.00	59.48	31.24	51.00	51.88	27.30	44.00
InfVar German	20.18	12.58	15.00	29.58	16.48	28.00	38.00	12.26	37.00
InfVar Turkish	31.86	13.33	28.50	36.39	13.08	31.00	30.76	12.36	29.00
PCC-A German*	88.64	8.78	91.17	79.49	15.05	80.92	75.76	9.98	75.97
PCC-A Turkish*	79.58	9.62	80.19	75.63	11.87	77.88	78.12	11.00	82.03

Note: **MoreG**: more language input in German, **Equal**: equal language input in both languages, **MoreT**: more language input in Turkish, *****: values presented correspond to %

On the contrary, more input in Turkish did not yield to better phonological performances on any Turkish outcome measures than more input in German. In fact, the average performance in these two categories was similar. Solely, experiencing a similar amount of input in both languages yielded to on average slightly weaker performances in Turkish outcome measures. However, standard deviations were large across measures and languages.

Statistical analyses using Kruskal-Wallis revealed that no outcome variable in Turkish differed significantly across the three language input categories ($H_{PhonVar}(2) = 1.80$, $p = .405$; $H_{InfVar}(2) = 2.42$, $p = .298$; $H_{PCC\text{-}A}(2) = 1.32$, $p = .516$). Phonological skills in German, however, were significantly affected by the proportion of input:

- Phonological variations: $H(2) = 14.40$, $p = .001$
- Infrequent variants: $H(2) = 12.79$, $p = .002$
- PCC-A-scores: $H(2) = 11.55$, $p = .003$.

Mann-Whitney tests including a Bonferroni correction (cut-off for significance $p < .017$) were used to follow up this finding (see Table 4.16).

Table 4.16: Between-group differences on phonological performance in German based on the relative amount of input

Outcome measure	MoreG – Equal				MoreG – MoreT				MoreT – Equal			
	U	***z***	***p***	***r***	***U***	***z***	***p***	***r***	***U***	***z***	***p***	***r***
PhonVar	225.50	-2.36	.017	-.32	52.00	-3.83	< .001	-.61	195.50	-1.74	.083	-.25
InfVar	243.00	-2.06	.039	-.28	64.00	-3.49	< .001	-.56	181.50	-2.03	.042	-.29
PCC-A	231.00	-2.27	.023	-.31	65.00	-3.46	< .001	-.55	219.50	-1.25	.216	-.18

Note: **MoreG**: more language input in German, **Equal**: equal language input in both languages, **MoreT**: more language input in Turkish, significant difference, approaching significant difference

Results of post-hoc analyses revealed that the relative amount of language input had a differential effect on German outcome measures when the language input categories MoreG and MoreT were compared. Children with an equal amount of language input did not differ significantly from those with more input in one of the languages on all German outcome measures but approached significance in the comparison of the MoreG vs. Equal categories for the number of PhonVar. For all significant differences effect sizes were large.

4.5.2.2 Differences within language input categories regarding outcome measures in German and Turkish

To investigate whether the phonological performances of children within each of the language input categories differed across languages Wilcoxon singed-rank tests were conducted. Results showed language-specific differences in the effect of proportion of input per day on phonological outcome measures in both languages (see Table 4.17).

Table 4.17: Within-group differences in phonological performance across languages based on the relative amount of input

Outcome measure	MoreG			Equal			MoreT		
	z	*p*	*r*	*z*	*p*	*r*	*z*	*p*	*r*
PhonVar GER vs. TUR	-2.96	.002	-.45	-.22	.828	-.03	-3.47	< .001	-.59
InfVar GER vs. TUR	-2.76	.004	-.42	-2.44	.013	-.30	-2.49	.011	-.43
PCC-A GER vs. TUR	-3.52	< .001	-.53	-2.39	.016	-.29	-1.92	.057	-.33

Note: **MoreG**: more language input in German, **Equal**: equal language input in both languages, **MoreT**: more language input in Turkish, significant difference, approaching significant difference

Children who experienced more input in a specific language mainly demonstrated significantly better phonological skills in this language over the other. However, children with an equal amount of input in both languages mainly showed significantly better phonological abilities in German than in Turkish. Effect sizes of significant findings were predominantly medium to large.

4.5.3 Effects of the proportion of language output on phonological performance

As for the proportion of language input, between-group and within-group differences were analysed to examine the effect of the proportion of language output on the children's phonological performances. Further, the same three types of categories as for input were used to refer to the proportion of language output (i.e., MoreG, Equal, and MoreT; see Section 3.2.3).

4.5.3.1 Differences between language output categories regarding phonological outcome measures

Similar to the presentation of the other influential factors, any age differences between the language output categories are reported first and results regarding output-related effects on phonological performances second.

Age differences in the language output variable

A Kruskal-Wallis test revealed that children in the different language output categories differed significantly on age, $H(2) = 6.543$, $p = .038$. Post-hoc analyses using Mann-Whitney tests and applying a Bonferroni correction (cut-off for significance $p < .017$), however, showed that none of the age differences across categories was statistically significant. The difference between the categories MoreG and Equal approached significance, though.

- MoreG ($Mdn = 56.00$) vs. Equal ($Mdn = 48.00$): $U = 217.00$, $z = -2.35$, $p = .018$, $r = -.32$
- MoreG ($Mdn = 56.00$) vs. MoreT ($Mdn = 47.00$): $U = 120.50$, $z = -1.98$, $p = .048$, $r = -.31$
- Equal ($Mdn = 48.00$) vs. MoreT ($Mdn = 47.00$): $U = 223.50$, $z = -.20$, $p = .846$, $r = -.03$.

Language output-related differences

As for input, proportionally more output in German resulted in on average better phonological performances on German outcome measures than an equal amount of output in both languages and more output in Turkish (see Table 4.18). However, more output in Turkish did not yield to better phonological performances on any Turkish outcome measures than more output in German. In fact, the average performances in these two categories were similar for the number of PhonVar and the PCC-A-scores. Experiencing a similar amount of output in both languages yielded to on average weaker performances on Turkish outcome measures, except for the number of InfVar. It should be noted, though, that standard deviations were large across measures and languages.

Table 4.18: Phonological outcome measures at t1 by proportion of language output

Outcome measure	MoreG (*n* = 24)			Equal (*n* = 29)			MoreT (*n* = 16)		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
PhonVar German	31.92	30.11	26.50	63.79	41.84	67.00	75.56	30.80	71.50
PhonVar Turkish	50.71	22.93	46.00	60.10	33.06	52.00	51.19	28.97	43.00
InfVar German	20.63	15.24	16.00	30.03	14.77	28.00	40.19	10.48	38.50
InfVar Turkish	34.46	14.25	29.00	35.38	12.39	31.00	30.75	13.01	28.50
PCC-A German*	88.52	10.54	91.52	77.89	14.35	78.09	75.95	9.71	77.57
PCC-A Turkish*	78.67	9.14	80.19	75.47	12.51	77.73	78.31	11.27	82.49

Note: **MoreG**: more language input in German, **Equal**: equal language input in both languages, **MoreT**: more language input in Turkish, *****: values presented correspond to %

A statistical analysis with the Kruskal-Wallis test revealed that no outcome variable in Turkish differed significantly across the three output proportion categories ($H_{PhonVar}(2) = 1.25$, $p = .534$; $H_{InfVar}(2) = 1.29$, $p = .526$; $H_{PCC\text{-}A}(2) = .89$, $p = .642$). Phonological skills in German, however, were significantly affected by the proportion of output:

- Phonological variations: $H(2) = 17.56$, $p < .001$
- Infrequent variants: $H(2) = 17.22$, $p < .001$
- PCC-A-scores: $H(2) = 14.35$, $p = .001$.

This finding was followed up with Mann-Whitney tests applying a Bonferroni correction (cut-off for significance $p < .017$). Results are presented in Table 4.19.

Table 4.19: Between-group differences on phonological performance in German based on the relative amount of output

Outcome measure	MoreG – Equal				MoreG – MoreT				MoreT – Equal			
	U	*z*	*p*	*r*	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*
PhonVar	172.00	-3.15	.001	-.43	50.00	-3.92	< .001	-.62	182.00	-1.19	.241	-.18
InfVar	211.50	-2.44	.014	-.34	50.00	-3.92	< .001	-.62	132.50	-2.36	.017	-.35
PCC-A	178.50	-3.03	.002	-.42	66.00	-3.48	< .001	-.55	208.00	-.57	.577	-.08

Note: **MoreG**: more language input in German, **Equal**: equal language input in both languages, **MoreT**: more language input in Turkish, significant difference, approaching significant difference

The data suggest that children demonstrated significantly better phonological skills on all German outcome measures when they proportionally spoke more German on an average day compared to both an equal amount of output in the two languages and relatively more output in Turkish. However, phonological performances did not differ significantly if the categories equal amount of output and MoreT were compared, except for the number of InfVar which approached statistical significance. For all significant differences effect sizes were moderate to large.

4.5.3.2 Differences within language output categories regarding outcome measures in German and Turkish

To explore possible language-specific differences in the effect of language output per day on phonological measures in both languages, Wilcoxon signed-rank tests were applied. The results clearly demonstrated that children who experienced more output in a specific language performed significantly better on phonological measures in this language, with PCC-A-scores in Turkish only approaching significance (see Table 4.20).

Differences in the phonological skills of children with an equal amount of output in both languages were not significant except for the number of InfVar which were fewer in German. The effect sizes of most significant differences were medium to large.

Table 4.20: Within-group differences in phonological performance across languages based on the relative amount of output

Outcome measure	MoreG			Equal			MoreT		
	z	*p*	*r*	*z*	*p*	*r*	*z*	*p*	*r*
PhonVar GER vs. TUR	-3.30	< .001	-.48	-.55	.594	-.07	-3.41	< .001	-.60
InfVar GER vs. TUR	-3.41	< .001	-.49	-2.07	.038	-.27	-2.73	.004	-.48
PCC-A GER vs. TUR	-3.80	< .001	-.55	-1.61	.110	-.21	-1.86	.065	-.33

Note: **MoreG**: more language input in German, **Equal**: equal language input in both languages, **MoreT**: more language input in Turkish, significant difference, approaching significant difference

4.5.4 Effects of language proficiency on phonological performance

To explore the effects of children's language proficiency in each language (as rated by their parents) on the phonological performances, between- and within-group differences were analysed. Between-group differences concern comparisons of the children's phonological performances in both languages between the different categories of

German/Turkish language proficiency, i.e., very good, good, satisfactory and poor proficiency (see Section 3.2.3). Within-group differences refer to comparisons of children's performances across languages within each of the different German/Turkish language proficiency categories.

4.5.4.1 Effects of language proficiency in German

As for all other influential factors, this section begins with a presentation of any age differences across language proficiency categories in German before reporting on the actual effects of language proficiency on children's phonological performances.

Age differences in the language proficiency variable

Statistical analyses using Kruskal-Wallis tests showed no significant age differences across the language proficiency categories in German, $H(2) = 4.273$, $p = .233$.

Differences between language proficiency categories in German regarding phonological outcome measures

The higher parents rated their children's language proficiency in German the better the children's phonological skills in German were on average across measures (see Table 4.21). For Turkish, however, this relationship was not as clear. Children with a very good proficiency in German also demonstrated on average the best phonological skills in Turkish. However, the average phonological performances in Turkish mainly remained similar across all other language proficiency categories in German.

Table 4.21: Phonological outcome measures at t1 by language proficiency in German

Outcome measure	Very good (*n* = 16)			Good (*n* = 30)			Satisfactory (*n* = 23)			Poor (*n* = 7)		
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*M*	*SD*	*Mdn*
PhonVar German	32.56	33.80	19.50	49.57	46.24	33.00	73.87	30.92	73.00	88.14	32.38	87.00
PhonVar Turkish	47.81	22.06	44.50	63.20	35.07	53.50	56.65	25.79	52.00	57.14	34.54	42.00
InfVar German	17.81	13.46	11.50	26.00	16.27	23.50	39.00	13.06	39.00	40.57	10.78	37.00
InfVar Turkish	33.25	12.16	30.00	37.03	13.32	35.00	33.78	13.82	29.00	33.57	16.02	30.00
PCC-A German*	89.09	11.63	93.29	82.75	15.67	88.87	75.31	10.41	75.97	71.93	9.63	74.91
PCC-A Turkish*	80.18	8.42	81.83	74.36	13.51	76.50	75.70	10.34	75.58	76.09	13.15	80.65

Note: *****: values presented correspond to %

To statistically explore the differences in the children's phonological performances when divided into the four German language proficiency categories Kruskal-Wallis tests were applied. Analyses supported the observations from the means (cf. Table 4.21) and revealed that no outcome variable in Turkish was affected by the children's language proficiency in German ($H_{PhonVar}(2) = 1.95$, $p = .584$; $H_{InfVar}(2) = 1.28$, $p = .735$; $H_{PCC\text{-}A}(2) = 2.25$, $p = .522$). Phonological skills in German, however, were significantly affected by the children's proficiency in this language:

- Phonological variations: $H(3) = 19.69$, $p < .001$
- Infrequent variants: $H(3) = 23.08$, $p < .001$
- PCC-A-scores: $H(3) = 17.71$, $p = .001$.

Post-hoc analyses using Mann-Whitney tests including a Bonferroni correction (cut-off for significance $p < .008$) were conducted to follow up this finding (see Table 4.22). Results revealed that not all language proficiency categories differed significantly from one another when phonological outcome measures in German were compared. The only three comparisons in which children's language proficiency did have a significant differential effect on phonological outcome measures were very good language proficiency versus satisfactory and poor language proficiency as well as good language proficiency versus satisfactory language proficiency. In all cases, better language proficiency resulted in better phonological skills. Effect sizes of all significant results were medium to large.

Table 4.22: Between-group differences on phonological outcome measures in German based on language proficiency in German

Outcome measure	Very good – Good				Very good – Satisfactory				Very good – Poor				Good – Satisfactory				Good – Poor				Satisfactory – Poor			
	U	***z***	***p***	***r***	***U***	***z***	***p***	***r***	***U***	***z***	***p***	***r***	***U***	***z***	***p***	***r***	***U***	***z***	***p***	***r***	***U***	***z***	***p***	***r***
PhonVar	165.50	-1.72	.087	-.25	60.50	-3.53	< .001	-.56	13.00	-2.87	.003	-.60	184.50	-2.88	.003	-.40	44.00	-2.37	.016	-.39	59.00	-1.06	.305	-.19
InfVar	157.50	-1.90	.057	-.28	49.50	-3.85	< .001	-.62	11.50	-2.98	.002	-.62	164.00	-3.25	.001	-.45	40.00	-2.52	.010	-.41	73.50	-.34	.746	-.06
PCC-A	163.00	-.178	.076	-.26	61.00	-3.51	< .001	-.56	15.00	-2.74	.004	-.57	198.00	-2.64	.008	-.36	53.00	-2.02	.043	-.33	64.00	-.81	.434	-.15

Note: significant differences, approaching significant difference

Table 4.23: Within-group differences in phonological performance across languages based on language proficiency in German

Outcome measure	Very good			Good			Satisfactory			Poor		
	z	***p***	***r***	***z***	***p***	***r***	***z***	***p***	***r***	***z***	***p***	***r***
PhonVar GER vs. TUR	-2.15	.030	-.38	-2.74	.005	-.35	-3.62	< .001	-.35	-2.37	.016	-.63
InfVar GER vs. TUR	-3.41	< .001	-.60	-3.29	.001	-.42	-2.13	.032	-.31	-1.15	.313	-.31
PCC-A GER vs. TUR	-2.84	.003	-.50	-3.96	< .001	-.51	-.34	.754	-.05	-2.03	.047	-.54

Note: **GER**: German, **TUR**: Turkish, significant differences

Differences within language proficiency categories in German regarding phonological outcome measures in German and Turkish

Exploring within-group differences per language, Wilcoxon signed-rank tests overall showed that children with very good and good language proficiency in German performed significantly better on all three German outcome measures than on Turkish measures (see Table 4.23). In contrast, children with satisfactory or poor language proficiency in German scored significantly better on most Turkish than on German outcome measures. For all significant differences effect sizes were medium to large.

4.5.4.2 Effects of language proficiency in Turkish

In this section, presentations of any age differences between the four language proficiency categories in Turkish precede the results for between- and within-group differences.

Age differences in the Turkish language proficiency variable

Analyses with the Kruskal-Wallis test revealed that children's age did not significantly differ across the four language proficiency categories in Turkish, $H(2) = .966$, $p = .810$.

Differences between language proficiency categories in Turkish regarding phonological outcome measures

Table 4.24 summarises the children's average phonological performance when divided by language proficiency in Turkish. The children's phonological skills in Turkish were weaker the lower their parents rated their language proficiency in this language. However, children with poor Turkish language proficiency stand out as their performance was on average either similar to or even better than that for children with satisfactory language proficiency. This finding is even more evident in the German phonological skills where children with poor proficiency in Turkish demonstrated on average the best scores on all German outcome measures. Standard deviations were high, though, and the number of children per category very imbalanced.

Table 4.24: Phonological outcome measures at t1 by language proficiency in Turkish

Outcome measure	Very good (*n* = 42)			Good (*n* = 27)			Satisfactory (*n* = 6)			Poor (*n* = 5)		
	M	***SD***	***Mdn***	***M***	***SD***	***Mdn***	***M***	***SD***	***Mdn***	***M***	***SD***	***Mdn***
PhonVar German	55.52	37.69	49.00	57.63	46.66	51.00	64.17	44.33	66.50	34.80	17.25	45.00
PhonVar Turkish	46.67	25.77	37.50	63.93	33.78	53.00	79.83	13.99	84.50	70.60	13.01	71.00
InfVar German	30.64	14.70	31.00	26.67	17.02	29.00	34.00	17.87	33.00	26.60	13.87	28.00
InfVar Turkish	29.62	11.61	26.50	36.85	11.13	32.00	48.67	13.38	49.50	47.00	12.31	46.00
PCC-A German*	81.09	12.56	81.10	80.47	15.29	82.33	77.79	15.68	77.56	86.36	4.91	84.45
PCC-A Turkish*	80.31	10.33	83.87	73.91	12.42	77.42	67.74	6.74	67.28	70.31	4.85	70.42

Note: *****: values presented correspond to %

Kruskal-Wallis tests were applied to statistically explore whether the phonological outcome variables differed significantly depending on the children's proficiency in Turkish. The analyses revealed that no outcome variable in German was affected by the children's language proficiency in Turkish ($H_{PhonVar}(2) = .51$, $p = .917$; $H_{InfVar}(2) = 1.87$, $p = .601$; $H_{PCC\text{-}A}(2) = .84$, $p = .839$). All phonological outcome measures in Turkish, however, were significantly affected by the children's proficiency in this language:

- Phonological variations: $H(3) = 14.50$, $p = .002$
- Infrequent variants: $H(3) = 17.51$, $p = .001$
- PCC-A-scores: $H(3) = 13.36$, $p = .004$.

This finding was followed up using Mann-Whitney tests and applying a Bonferroni correction (cut-off for significance $p < .008$). The results revealed that the number of PhonVar and PCC-A-scores were only significantly different when children reported to have very good language proficiency in Turkish were compared to those having satisfactory language proficiency. All other comparisons revealed no statistically significant differences in the outcome measures (see Table 4.25).

Regarding the number of InfVar in Turkish, a different picture could be found. Children who were reported to have very good language proficiency in Turkish produced significantly fewer InfVar in Turkish than those who were reported to have good, satisfactory and poor Turkish language proficiency. However, the number of InfVar in Turkish did not differ for other category comparisons. Effect sizes were medium.

Differences within language proficiency categories in Turkish regarding phonological outcome measures in German and Turkish

When comparing the children's performances across languages within each Turkish language proficiency group, statistical analyses applying Wilcoxon signed-rank tests demonstrated a less clear picture than for the influence of language proficiency in German (see Table 4.26). Children with very good proficiency in Turkish produced a significantly lower number of PhonVar in Turkish than in German. Their number of InfVar and their PCC-A-scores, however, did not significantly differ across Turkish and German. In contrast, children with good proficiency in Turkish showed no significant differences in the number of PhonVar in both languages but produced a significantly fewer number of InfVar and achieved higher PCC-A-scores in German than in Turkish. For those children, whose language proficiency in Turkish was rated as satisfactory or poor phonological outcome scores did not differ across languages but some differences were approaching significance.

Table 4.25: Between-group differences on phonological outcome measures in Turkish based on language proficiency in Turkish

Outcome measure	Very good – Good				Very good – Satisfactory				Very good – Poor				Good – Satisfactory				Good – Poor				Satisfactory – Poor			
	U	*z*	*p*	*r*	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*	*U*	*z*	*p*	*r*
PhonVar	355.50	-2.60	.009	-.31	35.00	-2.84	.003	-.41	44.00	-2.11	.033	-.31	53.00	-1.31	.201	-.23	50.50	-.88	.396	-.16	8.00	-1.28	.242	-.39
InfVar	343.00	-2.76	.005	-.33	38.50	-2.73	.004	-.39	30.50	-2.57	.007	-.38	59.00	-2.03	.041	-.18	37.50	-1.56	.125	.00	11.50	-.64	.571	-.19
PCC-A	371.00	-2.41	.015	-.29	39.50	-2.70	.005	-.39	43.00	-2.14	.030	-.31	51.00	-1.40	.169	-.24	49.50	-.93	.367	-.17	11.00	-.74	.535	-.22

Note: significant differences, approaching significant difference

Table 4.26: Within-group differences in phonological performance across languages based on language proficiency in Turkish

Outcome measure	Very good			Good			Satisfactory			Poor		
	z	*p*	*r*	*z*	*p*	*r*	*z*	*p*	*r*	*z*	*p*	*r*
PhonVar GER vs. TUR	-2.18	.029	-.24	-1.45	.152	-.20	-.94	.438	-.27	-1.75	.125	-.55
InfVar GER vs. TUR	-.55	.587	-.06	-2.90	.003	-.39	-2.02	.063	-.58	-2.02	.064	-.31
PCC-A GER vs. TUR	-.64	.528	-.07	-3.08	.001	-.42	-1.99	.063	-.58	-2.02	.063	-.64

Note: **GER**: German, **TUR**: Turkish, significant difference, approaching significant difference

4.5.5 Summary – Influential factors

The non-parametric tests applied here did not allow for a simultaneous controlling of age during analyses. Age differences between the individual categories were thus separately conducted and revealed significant differences for the language input categories with children in the MoreG category being the oldest. Further age differences could not be observed.

Analyses of potential influential factors revealed that gender seemed to have no or if only a marginal effect on children's phonological performances in Turkish and not at all on their phonological abilities in German regarding the three outcome variables (i.e., number of PhonVar, number of InfVar, PCC-A-scores).

The proportion of language input, however, was found to have a strong influence on the children's phonological performances in German. Proportionally more input in German compared to Turkish increased children's phonological performances across measures in German. In contrast, phonological performances in Turkish did not differ across language input categories. Nevertheless, language-specific effects of the proportion of language input could be found within categories. Children who received more input in German demonstrated more advanced phonological skills in this language compared to Turkish and vice versa. Children with an equal amount of input in both languages were found to have either comparable phonological skills in both languages or better skills in German on some phonological measures. Comparable language-specific effects were identified for the proportion of language output. Phonological performances within the category of equal language output, however, differed on less outcome measures across languages than for input.

Additionally, language proficiency had a language-specific influence on the three phonological outcome measures. Children with higher language proficiency in German performed better in German and weaker in Turkish and vice versa. In Turkish, however, this effect was predominantly obvious in the number of InfVar rather than in other measures.

5 DISCUSSION I – Phonological acquisition in Turkish-German bilingual children

The phonological acquisition in 84 Turkish-German bilingual children aged 3;0 - 5;5 years was analysed and factors influencing the speech acquisition were presented. This chapter aims to discuss the findings and answer the first two research questions:

1) How does the phonological acquisition in Turkish-German bilingual children proceed in both of their languages?
    a) At what rate do Turkish-German bilingual children acquire the consonants and consonant clusters of both of their languages?
    b) What type of phonological variations do Turkish-German bilingual children show in either/both of their languages?
2) Do age, gender, input and output patterns as well as language proficiency have an influence on children's phonological performances? And if so, in which way?

Although not explicitly listed, research question 4 and the related hypothesis (e), outlined Section 2.3, will be discussed and answered throughout all discussion chapters.

In order to describe the typical phonological development either longitudinal or cross-sectional studies are required (Bernhardt & Holdgrafer, 2001) and either spontaneous speech or single-word naming tasks can be used to examine the acquisition of consonants and CCs, the production of developmental phonological patterns and progressive quantitative performances (e.g., PCC-A, number of InfVar) (Bernhardt & Holdgrafer, 2001; Masterson et al., 2005; Wolk & Meisler, 1998). The collection of cross-sectional data has the benefit of testing large numbers of children across different age ranges in a comparably short time. As discussed in Section 2.1.1.1, single-word naming assessments have a further time-efficient advantage over the administration and evaluation of spontaneous speech data. However, for this the test tool needs to be thoroughly designed to allow for the identification of clinical markers by the respective measures in a satisfying way (Kirk & Vigeland, 2015; Messick, 1980). With this research project, such a tool is now also available for testing the phonological competences in Turkish of Turkish-German bilingual children (see Section 3.3.1.1). Thus, with this test it is now possible to collect cross-sectional as well as longitudinal data and provide information about the different aspects of phonological development as outlined above.

Still, there is the question of how much vocabulary knowledge (item familiarity) influences the results (see discussion in Section 5.1.2).

Comprehensive data on the typical phonological development in bilingual children are of twofold importance: (a) clinically they form a baseline for the identification of children with typical but also atypical development (Dodd, 1995; Stackhouse & Wells, 1997). This is important since atypically developing children are entitled to professional support in order to prevent later developmental difficulties (e.g., literacy acquisition issues). (b) These data add valuable insights to the theoretical discussion about the nature of phonological development in general and in bilingual children specifically (Babatsouli & Ingram, 2015; Davis & Bedore, 2013).

Hence, based on the results of the single-word naming tests administered at t1, this discussion chapter addresses four main aspects of the typical development in Turkish-German bilingual children:

a) Performance on quantitative measures (i.e., consonant and CC acquisition, PCC-A-scores and number of InfVar)
b) Use of phonological patterns
c) Production of cross-linguistic transfer phenomena and differences found across performances in German and Turkish
d) Influential factors on phonological acquisition.

Each of the sections will address performances in German and Turkish and include comparisons to existing monolingual and bilingual data. Those comparisons are restricted to the age of acquisition since the age of mastery was too sensitive to the small numbers in the present research.

5.1 Performances on quantitative measures

The basis for this discussion is formed by the results from the consonant and CC inventories as well as the PCC-A-scores and number of InfVar calculations in both languages at t1.

5.1.1 Rate of acquisition

Given the maturational aspect in speech acquisition (Davis & Bedore, 2013; McLeod, 2013; Weiss, 2007) it was hypothesised that cross-sectional data on the phonological acquisition in 3;0- to 5;5-year-old Turkish-German bilingual children would reveal a developmental trend across age groups. As expected, group analyses of the single-word naming performances revealed that the number of acquired consonants and CCs as well as children's PCC-A-scores increased with age, whereas their number of InfVar decreased across age groups. Full mastery of consonant and CC inventories, however, was not necessarily gained by children in the oldest age group. This developmental trend was evident in both languages although performances in German were slightly better across measures (e.g., more consonants and CCs acquired) than in Turkish. A general developmental trend in phonological acquisition across these measures was also reported for monolingual German- (Fox-Boyer, 2014b; Fox, 2016; Fox & Dodd, 1999; Schaefer & Fox-Boyer, 2016) and Turkish-speaking children (Ege, 2010; Topbaş & Yavaş, 2006) and is further commensurate with studies on bilingual children of other language combinations (e.g., Grech & Dodd, 2008; Lee et al., 2015b). This highlights the significant role chronological age or rather maturation (which is associated with chronological age) plays during both monolingual and bilingual speech acquisition.

Regarding the specific rate of acquisition, however, bilingual children were assumed to differ from monolingual children since they face the challenge of acquiring two phonological systems (more or less) at the same time. These systems are believed to interact during development which may have an effect on the rate and quality of the acquisition (Paradis & Genesee, 1996). If compared to monolingual children Turkish-German bilingual's rate of acquisition was, thus, hypothesised to show evidence for an accelerated, decelerated and similarly paced rate of acquisition.

Performances on quantitative measures in the present research revealed an overall slower phonological acquisition at the group level when retrospectively compared to those of age-matched German and Turkish monolingual children from earlier studies (see data presented in Section 1.4). This affected both languages and all quantitative measures included in this research (i.e., consonants, CCs, PCC-A-scores and number of InfVar). Two examples for this are children's PCC-A-scores and number of InfVar (i.e., phonological variations from adult-like speech which only occurred one to four times in the item set). Participants in the youngest age group of this research project (3;0 - 3;5 years) achieved PCC-A-scores in German of 67%, whereas monolinguals were reported to have achieved 74% phonemic accuracy already at the age of 1;6 - 1;11 years (Fox, 2000, 2007). Similarly, the Turkish-German participants produced a mean of 40 InfVar in German at the age of 3;0 - 3;5 years, whereas same-aged monolinguals produced on average 17.5 InfVar (Fox, 2016). In Turkish, the bilingual participants attained PCC-A-scores of 63.8% and 76.4% at the ages of 3;0 - 3;5 and 3;6 - 3;11 years, respectively, whereas monolinguals achieved PCC-scores of 94.85% at the age of 3;1 - 4;0 years (Topbaş & Yavaş, 2006). Data on InfVar in monolingual Turkish-speaking children do not exist for a comparison. Further differences in the rate of phonological acquisition in monolingual and bilingual children (e.g., consonant and CC acquisition) can be observed when comparing data presented in Table 1.3 - Table 1.7 (Sections 1.4.1 and 1.4.2 on monolingual acquisition) with those in Table 4.1 - Table 4.6 (Sections 4.1 and 4.2 on bilingual acquisition).

These differences between the rate of acquisition in monolinguals and bilinguals, however, were not restricted to the young age groups only but continued to be present until the age of 5;5 years when the bilinguals eventually approached monolingual competence in German (mean PCC-A-scores > 90%, mean number of InfVar 14.4). Their PCC-A-scores in Turkish, however, remained below those of monolinguals (82.1% vs. 97.52% on average). The difference between monolingual and bilingual children's consonant accuracy becomes even more obvious when the different measures used (i.e., PCC for Turkish monolinguals and PCC-A for bilinguals) are taken into account. The measure of PCC can be considered as "stricter" than the PCC-A as every phonetic deviation from the correct realisation of a sound is scored as an error (Shriberg & Kwiatkowsky, 1982b). In the PCC-A, however, common clinical phonetic deviations (Shriberg et al., 1997a) and phonetic CLT phenomena were accepted in this research (see Section 3.4.3).

Group findings, thus, suggest that bilingual children's "burden of acquiring two languages" may indeed interfere with the acquisition of their languages resulting in lagging behind monolinguals' pace as postulated with the deceleration hypothesis by Paradis and Genesee (1996, p. 4). However, it should be explored to what extent this hypothesis also holds true for individual performances since the questions arise as to whether all children perform slower than monolinguals, and if not, whether those who perform within the normal monolingual range may achieve scores above the monolingual average (i.e., evidence for the acceleration hypothesis).

Individual data of the present research revealed a large variability in participants' performances including some children who performed within the monolingual norm or even demonstrated an accelerated acquisition compared to monolinguals. This applied to both performances in German and Turkish. Sometimes evidence for an acceleration and deceleration occurred simultaneously within an individual. Two children in the 3;0 - 3;5-years age group, for example, had acquired all German consonants whereas monolinguals are reportedly still missing the consonant /ʃ/ at this age at the group level (Fox & Dodd, 1999). However, these children each had one coda CC (i.e., /çt/ and /lç/) missing from their inventories that monolinguals were reported to have already acquired by 2;5 years (Fox, 2004). Furthermore, eleven children (3;6 - 3;11: $n = 2$, 4;0 - 4;5: $n = 3$, 4;6 - 4;11: $n = 5$; 5;0 - 5;5: $n = 1$) were only missing the consonant /ʒ/ from their Turkish consonant inventories[12]. Although monolingual data suggest this sound is acquired by

[12]These were not necessarily the same children as those who met German monolingual norms.

3;0 years (Topbaş & Yavaş, 2006), Ege (2010) claims that it is even mispronounced in some adult speakers of Turkish due to its very infrequent occurrence. Thus, these children could be considered to demonstrate an acquisition rate for Turkish within the monolingual norm for this measure. Four of these eleven children (3;6 - 3;11 years: $n = 1$, 4;0 - 4;5 years: $n = 1$, 4;6 - 4;11 years: $n = 2$), however, only achieved PCC-A-scores between 85.71% and 88.48% in Turkish which are low compared to PCC-scores of 94.85% and 98.19% for 3;1 - 4;0- and 4;1 - 5;0-year-old Turkish monolinguals, respectively (Topbaş & Yavaş, 2006).

The implications of individual data are twofold. On the one hand, they confirm findings from the group results and support the deceleration hypothesis. On the other hand, they revealed a faster or similarly paced rate of acquisition as in monolinguals for some children supporting the (variation of) acceleration hypothesis (Fabiano-Smith & Goldstein, 2010b; Paradis & Genesee, 1996). The latter further illustrates that the deceleration found at the group level does not suggest a clinical delay in Turkish-German bilinguals' phonological acquisition but rather a slower acquisition than monolinguals at a given point of time in their phonological development (cf. Fabiano-Smith & Goldstein, 2010b). This is mainly supported by a comparison of children's rate of acquisition with that of other bilinguals. Participant's overall pace of consonant and CC acquisition was broadly consistent with that reported on Turkish-German speaking children in pilot studies (Salgert et al., 2012; Ünsal & Fox, 2002) as well as with the rate of phonemic acquisition in other bilingual children (Lee et al., 2015b; Mayr, Howells, & Lewis, 2015; Stow & Pert, 2006). Bilinguals in these studies also still had incomplete consonant and/or cluster inventories at the age of 4;6 - 5;5 years. Hence, needing more time to acquire the phonological systems of both languages compared to monolinguals does not seem to be unusual for bilingual children. The only exceptions are Turkish-German bilinguals' consonant accuracy scores as these were especially low in the younger age groups. Whereas Russian-English 3;3- to 5;7-year-old bilinguals achieved PCC-scores of 80% in English (Gildersleeve-Neumann & Wright, 2010), Spanish-English 4-year-olds achieved 94.1% in English (Goldstein & Washington, 2001), and Maltese-English 3;0- to 3;5-year-olds' average PCC-scores were 93.9% in Maltese (Grech & Dodd, 2008) Turkish-German bilingual's did not achieve PCC-A-scores larger than 80% before the age of 4;6 in German and 5;0 years in Turkish. Possible explanations for these differences may be the different sample sizes in the studies which can have a significant effect on percentages and mean values as well as Turkish-German bilingual's low familiarity with the test items of the TPA and PLAKSS-II (see Section 5.1.2 for a further discussion on this).

Overall, outcomes of group analyses confirmed findings from other studies in which bilinguals' rate of acquisition was considered to lag behind that of monolinguals' (Bunta et al., 2009; Fabiano-Smith & Goldstein, 2010b; Gildersleeve-Neumann et al., 2008; Goldstein & Washington, 2001; Holm & Dodd, 2006). Also, the simultaneous presence of a slower, faster and similarly paced rate of acquisition as in monolinguals found in individual data supports findings from studies with other bilingual cohorts (e.g., Fabiano-Smith & Goldstein, 2010b; So & Leung, 2006). Taken together, the occurrence of these differences in the acquisition rates suggest that a bilingual child's languages interact during phonological development (Paradis & Genesee, 1996). To what extent and with what frequency this interaction may occur will be discussed in Section 5.3.1.

Moreover, these differences in the rate of acquisition compared to monolinguals also seem to support the hypotheses of positive and negative transfer within the Unified Competition Model (MacWhinney, 2005a). However, the premises behind MacWhinney's (2005a) model are fundamentally different to Paradis and Genesee's (1996) hypothesis in that MacWhinney (2005a) supports a functionalist rather than a formalist approach and gives weight to the input children receive. In his view, the fact that bilingual children's input is divided across two languages has an effect on which linguistic aspects are likely to form frequent and strongly reliable cues for the children (i.e., shared aspects between the languages = frequent and strongly reliable cues, unshared aspects = less frequent and reliable cues) which in turn either accelerate or decelerate their rate of acquisition. Hence, to fully confirm his hypothesis it is necessary to explore whether there is evidence for these assumptions in the acquisition rates of shared versus unshared aspects. For this, the rates of acquisition of shared and unshared consonants between German and Turkish as well as that of the shared syllable structure of coda CCs and of the unshared syllable structure of onset CCs were individually compared to the respective rates of acquisition in monolinguals.

This comparison, however, neither revealed a faster acquisition of shared consonants nor of coda CCs in both languages compared to monolinguals. But, the majority of shared consonants (i.e., /m, n, p, b, t, d, k, f, s, l, j/ in German and /m, p, b, t, d, k, f, s, j/ in Turkish) as well as a few coda clusters (i.e., /ŋk/, /nt/, and /lt/ in German; /nk/ in Turkish) were already acquired by the youngest age group (3;0 - 3;5 years). This could demonstrate an acquisition rate within the monolingual norm (i.e., positive transfer) but since most of these consonants and coda clusters are already acquired by monolingual children under the age of three, data from younger bilingual children would be needed to determine the exact age of their acquisition. Despite no clear evidence for positive transfer, there was definite evidence for negative transfer since none of the unshared consonants and German onset CCs were acquired within the monolingual norm.

Hence, the data from the present research can only partly support MacWhinney's (2005a) model and suggest that similarities and differences between languages may not be the only factor accounting for differences/similarities between monolingual and bilingual children's rate of acquisition. Besides personal factors such as individual learning strategies, processing abilities and maturation paces (Davis & Bedore, 2013; Pierrehumbert, 2003; Stoel-Gammon & Dunn, 1985; Vihman, 2014) that may affect an individual's rate of acquisition, children's experience patterns with their two languages come into consideration. These are likely to differ significantly within and between individuals so that effects of the frequency and the amount of input and output in a language should be controlled for/investigated (Goldstein & Bunta, 2012). For this reason, the effect of the proportion of language input and output as well as language proficiency on children's phonological performances was investigated within this research project. The outcomes of these analyses are discussed in Section 5.4. Additionally, methodological aspects may have confounded the results (see Section 5.1.2).

Overall the formulated hypothesis within this project regarding the differences in children's rate of acquisition compared to monolingual children could be confirmed. The rates differed noticeably and demonstrated an interaction between the bilingual's two languages (cf. Paradis & Genesee, 1996). However, although it is frequently conducted, the comparison of bilingual acquisition rates with those by monolinguals has an important disadvantage. It encompasses the evaluation of only one language at a time but does not take into account the children's acquisition and achievements in their other language. Thus, it needs to be questioned whether these comparisons to monolinguals are actually useful and trustworthy if they are conducted in this way. Rather, it may be necessary for future investigations to adduce the rate of acquisition of phonological features in both languages of a bilingual. For example, if investigating the rate of consonant acquisition, it may be more representative to add the number of consonants acquired in both languages to the number of all consonants acquired for only one language and then compare the sum of these to the number of consonants acquired in same-aged monolinguals of each language, as it is recommended for the evaluation of vocabulary size (e.g., Scharff Rethfeld, 2013). This would then include the whole phonological knowledge of a bilingual achieved at a given point in time and illustrate their understanding of language-specific and shared features.

5.1.2 Possible confounding variables and aspects

As pointed out throughout the previous sections, there might be further explanations for why children's performances appeared to be weaker/slower than monolinguals'. One possible confounding factor is the overall sample size in this research ($N = 84$). Although it is comparably large for internationally published data on bilingual children's phonological acquisition (see Hambly et al., 2013 for an overview), the numbers of participants in each age group are very heterogeneous with some being very small (i.e., 3;0 - 3;5 years: $n = 12$, 5;0 - 5;5 years: $n = 7$). This limits the overall representativeness and generalisability of group values (especially means and percentages) and had an effect on consonant and CC inventory analyses in the present research. Due to the commonly used but still arbitrary cut-off values of 75% (acquisition) and 90% (mastery) for these analyses (e.g., Dodd, 1995; Fox & Dodd, 1999; Topbaş & Yavaş, 2006; Zhu, 2000), one more child or one less child in an age group could have affected the age of acquisition/mastery for an age group. One example for this is the acquisition of the German onset CC /kn/ which was not considered to be acquired by the oldest age group since two of the seven children did not produce the cluster accurately which in turn affected the 75% acquisition threshold. Hence, this might be an explanation for the dissimilar rates of acquisition in monolinguals and bilinguals as well as across German and Turkish. Sample size issues, therefore, need to be considered when interpreting data from the present research but also from other referenced work given that there were not many projects with larger sample sizes.

Further, it needs to be acknowledged that there were methodological differences in the referenced studies which concerned the criteria employed for considering a consonant/CC to be acquired. For comparisons with monolingual data these differences also affected the use of different assessment material (except for some data in Fox-Boyer, 2014b; Fox-Boyer and Schaefer, 2015; Schaefer & Fox-Boyer, 2016). This limited the comparability to findings of the present research and may have also resulted in differences across studies.

In addition to the aspects outlined above, familiarity issues with the items in the Turkish and German phonology assessment need to be acknowledged. Although items of the PLAKSS-II were assumed to be known by the children and items for the TPA were very carefully selected (see Section 3.3.1.1), analyses in Section 3.3.1.3 revealed that children could not name all items of the TPA and the PLAKSS-II spontaneously. Their familiarity with TPA-items was noticeably lower than with PLAKSS-II-items. Given the short time of contact to German (i.e., 7 months) for some children and these item familiarity/vocabulary issues (especially in Turkish), it remains unclear how stable and distinct children's phonological representations and motor programs for the assessed words actually were (cf. Macrae, Tyler, & Lewis, 2014). It is conceivable that children had to depend more on their abilities to assemble new motor programmes for unknown words (Stackhouse & Wells, 1997) than on activating their existing motor programs for known words. The weaker performances and slower acquisition rates than in monolinguals may, therefore, in fact be influenced by word knowledge rather than the bilingual children's additional processing and memory load. Especially the remarkably large number of InfVar and low PCC-A-scores children demonstrated in both of their languages seems to reflect their variability in production, which can especially occur if unfamiliar words are produced for the first time (Stackhouse & Wells, 1997). Further, the analyses of children's CC inventories may have been biased by this since the acquisition of a CC was based on the first occurrence of a cluster in the assessment (i.e., only on a single occurrence of a CC), thus the degree of familiarity with a test item was of increased significance. These findings support the claim of several researchers that children's familiarity with the items in a test tool is vital to collect representative data on their phonological acquisition (Eisenberg & Hitchcock, 2010; James, 2001b; Stoel-Gammon, 2011; Stoel-Gammon & Williams, 2013). However, there are at least two issues with estimating bilingual children's vocabulary skills. First, bilingual children's vocabulary skills are highly influenced by the frequency and amount of language input (Parra et al., 2011; Pearson et al., 1997) which is known to vary across bilingual speakers. Second, a bilingual's lexicon is distributed over two languages including words known in one language only and those known in both languages (Peña et al., 2002). Thus, to rule out any bias by children's vocabulary knowledge, it may be necessary to assess children's vocabulary skills in future research as well.

Moreover, the syllable position in which a consonant was assessed may have confounded the rate of acquisition. Children's apparent later acquisition of the consonant /g/ in German, for example, is likely to be ascribable to their difficulty to acquire /g/ in unstressed syllable-positions preceding a stressed syllable. This was the syllable-position in which /g/ was assessed the most (5 out of 11 times) in the PLAKSS-II. Similar to monolingual German-speaking children (Fox, 2016), Turkish-German bilinguals frequently deleted the whole (weak) syllable in which /g/ occurred, which hindered the children to achieve the cut-off for acquisition. It may therefore seem recommendable for future evaluations to analyse the acquisition of /g/ separately for those in pre-stressed positions and those in all other syllable positions to identify whether children's difficulties are with the phoneme /g/ or the syllable position.

Finally, the quality of input children receive through their interlocutors, especially the phonetic-phonological quality, is unknown for this research but may have impacted on their phonological performances (cf. Holm et al., 1999; Khattab, 2006). It is assumed that Turkish-German bilingual children come into contact with a number of dialectal, accentual and ethnolectal varieties of German and Turkish which may have formed their phonological representations and/or output forms. Despite paying attention to this aspect in the present research by not scoring dialectally, accentually and ethnolectally influenced speech as incorrect (see Section 3.4) it is still possible that not all characteristics of Turkish German and German Turkish have been considered due to the lack of knowledge and studies in this field (see Sections 1.3.2.4 and 1.3.3). This also applies to common conversational speech processes in German and Turkish. Children's frequent realisation of the /ft/-cluster as [f] in Turkish, for example, corroborates conversational speech processes in German where /t/ in syllable-/word-final CC is often deleted (Zimmerer, Scharinger, & Reetz, 2011) because speakers' speed of pronunciation hinders them to reach the correct articulatory position for /t/ in time (Lindblom, 1990). This realisation pattern of /ft/ was also frequently observed in the speech of the Turkish-German test assistants. Hence, participants and/or their interlocutors may have applied the common conversational speech pattern in German to their productions in Turkish (if it was not present in this language anyway).

5.2 Phonological patterns produced by Turkish-German bilingual children

In addition to the quantitative aspects discussed in the previous sections, this section addresses the qualitative aspects of children's level of speech inaccuracy by discussing the phonological patterns (i.e., phonological variations that occurred at least five times in the data of an individual) observed in the children's speech with regard to the current literature on monolingual and bilingual children.

Based on the literature on monolingual and bilingual children it was expected that bilingual children demonstrate a variety of structural and systemic simplifications in both of their languages which show similarities but also differences to monolingual children due to an interaction of the two linguistic systems (Holm & Dodd, 1999c). In order to explore this for the population of Turkish-German bilingual children, the different types of phonological patterns produced by the children in German and Turkish were analysed.

Data of the present research revealed that children exhibited 14 different structural and systemic phonological patterns in both languages (see Table 5.1). As expected by maturation effects (Davis & Bedore, 2013; McLeod, 2013; Weiss, 2007) and analogous to an age-related increase in PCC-A-scores, consonants and CCs acquired (see Section 5.1.1), the number of phonological patterns present per age group decreased with age. Hence, some patterns occurred across all age groups (e.g., fronting of post-alveolar sibilants) and others only occurred in the younger age groups (e.g., metathesis). The greatest variability in pattern types is found in the three-year-olds. Hence, the bilingual cohort in this research generally demonstrates a developmental trend in achieving phonological accuracy.

Table 5.1: Comparison of the presence of phonological patterns observed in this research with their presence in monolingual German- and Turkish-speaking children and smaller cohorts of Turkish-German bilingual children

Lang.	Phonological patterns	Occurrence in ML children of same language[1]	Occurrence in ML children of respective other language[2]	Presence in other Turkish-German BL children[3]
German	Initial CC reduction	+	NA	+
	Weak syllable deletion	+	+	+
	Final CC reduction	+	+	+
	Intrusive consonants	-	-	-
	Fronting of /ʃ/	+	+	+
	Deaffrication	+	+	+
	Devoicing	+	word-final only	-
	Stopping	+	+	-
	Assimilation	+	+	-
	Fronting of /k, g, ŋ/	+	rare	-
	Backing of /t, d, n/	-	+	-
	Metathesis	-	+	-
	Voicing	+	word-initial only	-
	Labialisation	-	-	-
Turkish	SF consonant deletion	+	+	+
	Final consonant deletion	+	only /l/	-
	Initial consonant deletion	-	-	-
	Final CC reduction	+	+	-
	Intrusive consonant	-	-	-
	Fronting of /ʃ, ʒ, t͡ʃ, d͡ʒ/	+	+	+
	Devoicing	word-final only	+	+
	Liquid deviation /ɾ/ → [l]	+	NA	+
	Assimilation	+	+	+
	Deaffrication	+	+	+
	Gliding of /ɾ/	+	NA	-
	Stopping	+	+	+
	Metathesis	+	-	-
	Backing of /s, z/	+	-	-

Note: **ML**: monolingual, **BL**: bilingual, [1]Monolingual data come from Fox-Boyer (2014b) and Fox-Boyer and Schäfer (2015) for German and Topbaş and Yavaş (2006) for Turkish, [2]Phonological patterns present in the Turkish-German bilingual's German productions are compared to phonological patterns present in Turkish-speaking monolinguals and vice versa to identify possible transfer effects. [3]Bilingual data come from pilot studies by Ünsal and Fox (2002) and Salgert et al. (2012), **SF**: syllable-final, **patterns identical in German and Turkish productions**, structural simplifications, unshaded: systemic simplifications, **+**: pattern occurs in the typical development of this language, **-**: pattern does not occur in the typical development of this language, **NA**: pattern not applicable for this language

As hypothesised above, Turkish-German bilinguals were found to also produce a large number of patterns in both languages congruent with typically developing German- ($n = 10$) and Turkish-speaking ($n = 13$) monolinguals (see Table 5.1). All patterns except initial consonant deletion, intrusive consonant, labialisation and backing of /s, z/ have frequently been reported to occur in children's speech cross-linguistically (McLeod, 2007; Zhu & Dodd, 2006). Thus, the presence of these patterns in both bilingual and monolingual children's speech may suggest that children generally tend to use similar simplification strategies across languages to cope with more complex phonological structures (Fox, 2004). This would support the view of language universals guiding speech acquisition held by most formalist approaches (see Section 1.1). However, this was not the case for all patterns (see those indicated with a "-" in Table 5.1) and group as well as individual data revealed that phonological patterns were not identical in children's German and Turkish productions (see patterns highlighted in red and bold in Table 5.1). This means some patterns for which phonotactic conditions would have allowed an occurrence in both languages were only present in one (e.g., fronting of /k, g/ in German and final consonant deletion in Turkish). This suggests language-specific acquisition strategies, which in turn imply that language universals cannot fully explain the types of patterns that occurred in Turkish-German bilingual's speech. Rather, influences of ambient language phonology (e.g., frequency of occurrence of phonological structures, phonological saliency) and an interaction between the two within the bilingual speaker (both further discussed in Section 5.3) as well as children's input and output patterns (discussed in Section 5.4.2) come into consideration.

Moreover, the timely occurrence of patterns mutual with monolinguals often differed for the respective populations. Monolingual German- and Turkish-speaking children, for example, stopped using deaffrication by the age of 2;5 years (see Sections 1.4.1 and 1.4.2) and were generally not reported to show any phonological pattern beyond the age of 4;11 (Fox-Boyer, 2014b, 2014c; Fox-Boyer & Schäfer, 2015; Topbaş & Yavaş, 2006). In contrast, Turkish-German bilinguals exhibited deaffrication until the ages of 4;0 - 4;5 (German) and 4;6 - 4;11 (Turkish) and generally produced phonological patterns until the oldest age group (5;0 - 5;5 years). These differences in the age of overcoming phonological patterns could be observed for five patterns in German and ten in Turkish whereby the exact time differences varied per pattern. This suggests an overall tendency for a decelerated acquisition compared to monolinguals which was also found in quantitative measures (see discussion in Section 5.1.1). However, there was also evidence for an accelerated and similar rate of overcoming phonological patterns compared to monolinguals. Weak syllable deletion in German and final CC reduction in Turkish, for instance, were overcome earlier than by the respective monolingual cohorts (Fox-Boyer, 2014b; Topbaş & Yavaş, 2006) and final CC reduction and fronting of /k, g/ in German were overcome at a similar age as in German-speaking monolinguals (Fox-Boyer, 2014b). The simultaneous presence of (a variation of) an accelerated and decelerated phonological acquisition compared to monolinguals confirm outcomes reported for other bilinguals (Fabiano-Smith & Goldstein, 2010b; So & Leung, 2006) and provide support for the Interdependence Hypothesis (Paradis & Genesee, 1996) and the Unified Competition Model (MacWhinney, 2005a). Thus, the first hypothesis (i.e., regarding a difference in the rate of acquisition) is also confirmed for qualitative data.

A comparison of the observed phonological patterns in the present research to those reported for similarly-aged typically developing Turkish-German bilinguals in pilot studies (Salgert et al., 2012; Ünsal & Fox, 2002) showed some agreements as well. The number of mutual patterns was comparably small which may be ascribable to the very different sample sizes (i.e., *n* = 19 and *n* = 20 vs. *n* = 84). However, agreement was on those patterns that were found to occur relatively frequently in the present sample. Therefore, findings from smaller cohorts could be replicated and enhance the probability that the observed phonological patterns are generalisable to the larger population of Turkish-German bilingual speakers. Reversely, some patterns found in the smaller cohorts' productions could not be observed in at least 15% of any age group in the present research (i.e., final consonant deletion and fronting of /ç/ in German; strong syllable deletion, affrication and voicing in Turkish; Salgert et al., 2012; Ünsal & Fox, 2002) which may indicate that these are rather unusual for typically developing Turkish-German bilinguals.

The question of whether patterns are truly typical for a respective age group of a given population is crucial for two reasons. First, knowing the typical phonological patterns for a certain population provides theoretical insights into children's acquisition strategies (Vihman, 2014). Second, it helps to distinguish atypically from typically developing children in a clinical setting (Dodd, 1995). Therefore, it is internationally required that normative data are representative of the skills in question and that their sample sizes allow for generalisability to a larger population (Kirk & Vigeland, 2014; see Section 2.1.1.2). In this regard, the frequency of occurrence of a phonological pattern in an age group as well as the frequency to which it is produced by an individual are considered crucial cut-off values for the phonological pattern analysis (Dodd, 1995; Zhu & Dodd, 2006). Considering the specific cut-offs applied in this research (i.e., 5x within an individual and at least within 15% of an age group; see Sections 3.4.4 and 3.4.8.1) as well as the overall sample size, the question arose of what extent the types of patterns observed in the present cohort (see Table 5.1) were likely to be representative for the typical phonological acquisition in 3;0- to 5;5-year-old Turkish-German bilingual children.

When looking at the frequency of occurrence of the identified patterns two groups of patterns could be identified: (a) patterns that fulfilled both frequency of occurrence criteria easily (i.e., pattern occurred significantly more than 5x in an individual and in significantly more than 15% of an age group) and for more than one age group, and (b) patterns that only just met the cut-off values for these criteria and were only present in one (or a maximum of two) age group(s). Due to their frequent occurrence within and across children, patterns in group (a) were in fact anticipated to be typical for the Turkish-German bilingual cohort. These included initial and final CC reduction, weak syllable deletion, fronting /ʃ/, deaffrication, devoicing, stopping and assimilation in German and syllable-final consonant deletion, fronting /ʃ, ʒ, t͡ʃ, d͡ʒ/, devoicing, liquid deviation, assimilation, deaffrication, gliding /r/, and stopping in Turkish. The fact that all of these patterns were also reported to occur in typically developing German-speaking (Fox-Boyer, 2014b, 2014c; Fox-Boyer & Schäfer, 2015) and Turkish-speaking monolingual children (Topbaş, 2006; Topbaş & Yavaş, 2006) further supports this assumption. It has to be noted though, that the devoicing pattern has been excluded from more recent analyses in German-speaking monolinguals since it is part of a range of regional variations of German (Barbour & Stevenson, 1990; Fox-Boyer, 2014c; Fox-Boyer & Schäfer, 2015; Wiese, 2011). Hence, its occurrence in children's productions could be due to either dialectal/accentual or developmental issues or a combination of both (cf. Kim et al., 2016). In Turkish, devoicing is only considered as typical if affecting word-final positions (Topbaş & Yavaş, 2006). Children in the present research, however, most frequently devoiced the syllable-final fricatives /v/ and /z/ as well as word-initial /d͜ʒ/. Thus, those instances could be assumed to be due to an overgeneralisation of the German rule of word-final devoicing to other syllable positions (cf. Holm & Dodd, 1999c).

For patterns that were allocated to group (b), in contrast, representativeness and generalisability remain questionable. The respective patterns (i.e., intrusive consonants, fronting /k, g, ŋ/, voicing, backing /t, d, n/, metathesis and labialisation in German and initial and final consonant deletion, final CC reduction, intrusive consonant, metathesis and backing /s, z/ in Turkish) only occurred in a small percentage of children (17 - 33%) in the youngest age group (3;0 - 3;5 years) and at a low frequency (i.e., just around the cut-off of five tokens) per participant. About half of them are additionally considered deviant for monolingual children (Fox-Boyer, 2014a; Topbaş, 2006) and are cross-linguistically only rarely reported (e.g., intrusive consonants in Australian-English speaking children (James, 2001a; here described as epentheses of consonants), initial consonant deletion in Finnish- and Thai-speaking children (Lorwatanapongsa & Maroonroge, 2007; Savinainen-Makkonen, 2000), and labialisation in Thai-speaking children (Lorwatanapongsa & Maroonroge, 2007)). Their presence in the Turkish-German bilinguals of the present research may therefore be subject to chance caused by the small sample size and the arbitrary cut-offs used. Alternatively, some patterns may have been transferred from the other language (e.g., metathesis, backing) for which they are considered typical (see Table 5.1), occurred due to a reorganisation of children's phonological systems triggered by differences between German and Turkish phonological systems (Holm & Dodd, 1999c), or may be due to an overgeneralisation of newly acquired phones (Menn, 1981; Vihman, 2014). For the categorisation process in the present research (see Section 8.1), however, all patterns occurring in at least 15% of an age group were considered to be typical for Turkish-German bilingual children for reasons outlined in Section 3.4.8.1.

Overall, the presence of these two groups emphasises how sensitive 'normative' data are for sample size, since the number of participants included in a sample may not only have an effect on the information they provide regarding typical development but also on the evaluation of what type of behaviour is considered to be atypical. Further, the frequency of occurrence of a pattern in a population is strongly dependent on the sample tested and the cut-off for the frequency of occurrence used (Kirk & Vigeland, 2015; McReynolds & Elbert, 1981), which emphasises children's unique learning and acquisition strategies (Davis & Bedore, 2013; Vihman, 2016). Consequently, the use of five necessary occurrences, as used in this research and other studies (Dodd et al., 2003; Fox-Boyer, 2014a; Holm & Dodd, 1999a, 1999c; Holm et al., 1997; Kim et al., 2016), may either under- or over-represent the presence of phonological patterns in children's speech. Future investigations with larger sample sizes but also a follow-up assessment (see Section 7.2 for a discussion of the findings) will shed light on the degree of transience of these 'uncommon' patterns in Turkish-German bilingual children. This will be particularly helpful for the clinical decision-making process, since especially backing of alveolar obstruents and initial consonant deletion are considered to be one of the crucial red flags for children with suspected SSD in many languages (e.g. Cheung & Abberton, 2000; Fox-Boyer, 2014a; Shriberg et al., 2003; Zhu & Dodd, 2000a).

5.3 Cross-linguistic transfer phenomena and comparison of Turkish-German bilingual children's two languages

Two characteristics of bilingual speech are the occurrence of CLT and distributed skills in children's two languages (see Sections 1.2.3.1 and 1.2.3.3). Both aspects are of clinical importance when assessing bilingual children since the misidentification of CLT as atypical phonological patterns and the examination of only one of a bilingual child's two languages may lead to an incorrect clinical diagnosis (e.g., McLeod et al., 2017; Yavaş & Goldstein, 1998), which is to be avoided. Hence, both aspects were explored for 3;0- to 5;5-year-old Turkish-German bilingual children in the present research and the results are discussed in the following two subsections.

5.3.1 Cross-linguistic transfer phenomena

Models and hypotheses on bilingual children's language acquisition usually include the aspect of cross-linguistic transfer (Grosjean, 2013; MacWhinney, 2005a, 2005b; Paradis & Genesee, 1996) since researchers assume an at least temporarily interaction of a bilingual's two languages (Paradis & Genesee, 1996). Given the frequent reports on the occurrence of CLT phenomena in bilingual children's speech productions (e.g., Hambly et al., 2013) it was hypothesised that Turkish-German bilingual children would also transfer linguistic features and rules from one language to the other. These were expected to occur bi-directionally within the sample but be rather infrequent. To identify possible occurrences of CLT, the phonologies of Turkish and German were contrasted systematically and children's phonological variations analysed qualitatively for any instances of segmental and suprasegmental transfer.

As expected, children in the present research produced types of phonetic and phonological variations in both of their languages that could be explained by CLT. These affected the segmental and suprasegmental level, were produced across age groups but generally occurred infrequently (i.e., in less than 30% of an age group). Hence, the formulated hypothesis (i.e., the infrequent and bi-directional occurrence of transfer) could be confirmed.

The most prominent CLT phenomena were substitutions of the /r/-phone (e.g., /ʁ/ → [ɾ, r, l, j] in German and /ɾ/ → [ʁ, ɐ] in Turkish). These were assumed to be a compensation of the inability to produce the target sound by relying on similar (sounding) phones/structures from the other (probably more proficient) language. The suprasegmental transfer of the Turkish developmental pattern /ɾ/ → [l, j] to German, for example, usually only occurred when the children had not acquired the Turkish [ɾ] yet (also cf. Salgert et al., 2012). It is hypothesised that the children were intending to transfer the Turkish-specific /r/-phone to their German productions, but since they were not able to produce this phone accurately (e.g., due to motoric difficulties), they applied the same (phonological) simplification pattern as in their L1. This principle may also apply to other systemic simplification patterns children transferred from the non-target languages (e.g., metathesis in German; see Section 5.2). However, since these are considered atypical for the target language, CLT effects and atypical development are difficult to differentiate in those cases. Other transfer occurrences may be the result of a confusion by differences in the complexities of the respective phonological systems (e.g., the use of a German stress pattern in Turkish; cf. Gildersleeve-Neumann et al., 2009) or reflect the coping with unfamiliar/less practised syllable structures (i.e., vowel epenthesis in German initial CCs; cf., Yavaş, 1998).

All of the observed types of CLT were either also reported in previous studies on Turkish-German bilingual children's speech acquisition (Salgert et al., 2012; Ünsal & Fox, 2002) or can be ascribed to *Türkendeutsch* – a pidginised version of the German used by Turkish speakers in Germany (Tekin, 2011; Tekin & Colliander, 2010). Hence, their presence in Turkish-German bilingual children's speech is likely to be common and may either be due to a mutual influence of German and Turkish phonologies during children's development (Yavaş & Goldstein, 1998) or be a reflection of the phonologies in the children's input (i.e., varieties of German and Turkish which differ from the mainstream variety of pronunciation) (Khattab, 2006; McLeod, 2007). To clearly differentiate between the two, future investigations should ideally include a phonetic-phonological analysis of the speech of children's main interlocutors (e.g., parents, siblings, nursery teachers) and compare these with children's productions.

Despite CLT occurring bi-directionally in the present data, there was a tendency for more transfer occurrences in German than in Turkish productions (see Section 4.3.4). Keshavarz and Ingram (2002) suggest that CLT occurrences are closely linked to language dominance. That is, language exposure and usage patterns guide the direction and frequency of transfer, in that sounds from the dominant language are more frequently transferred to the non-dominant language than vice versa. Thus, the type and degree of CLT can vary from one bilingual speaker to another. Taking on this suggestion would imply that there were more Turkish-dominant than German-dominant participants in the sample. This, however, would warrant further investigation since language dominance and CLT were not included as variables in the evaluation of influential factors within this thesis and the number of children showing CLTs were relatively small.

Further, CLT only occurred in a small number of Turkish-German bilingual children and mainly to a low frequency within a child. This is in line with studies on other language combinations (Fabiano-Smith & Goldstein, 2010b; Fabiano & Goldstein, 2005; Gildersleeve-Neumann et al., 2008; Gildersleeve-Neumann et al., 2009; Goldstein & Bunta, 2012; Goldstein et al., 2005; Keshavarz & Ingram, 2002; Schnitzer & Krasinski, 1994, 1996) and suggests that the interaction between German and Turkish is rather limited (Fabiano-Smith & Goldstein, 2010b; Paradis, 2001). It also emphasises that most children keep their language systems separate (Fabiano-Smith & Goldstein, 2010b). The correct identification of CLT, however, is crucial for the clinical decision-making process to prevent misinterpretations of phonological patterns (see Section 2.1.2 for a further discussion). Hence, the analysis of CLT should not be neglected or undervalued.

5.3.2 Comparison of performances across languages

Since bilingual children have very different experiences with their two languages (e.g., input and output patterns, context dependent exposure), they cannot be considered as a monolingual person times two (Grosjean, 1989). Therefore, it was anticipated that children's performances might differ across their German and Turkish single-word naming productions. Data of the present research revealed that across all quantitative and qualitative measures, children indeed demonstrated distributed skills in their two languages which could be observed in both group and individual data. These expressed themselves in form of:

- a different number of consonants acquired in German and Turkish
- dissimilar acquisition rates for some shared sounds across languages
- different PCC-A-scores and numbers of PhonVar and InfVar in the two languages
- diverse phonological patterns in German and Turkish
- the use of phonological patterns in only one language that could have occurred in both languages
- dissimilar ages of overcoming phonological patterns mutual to German and Turkish.

Imbalances in the phonological skills across languages have been reported for various bilingual populations (e.g., Anderson, 2004; Brice et al., 2009; Bunta et al., 2009; Holm & Dodd, 1999c; Holm & Dodd, 2006; Holm et al., 1999; Kim et al., 2016; Yang & Zhu, 2010) and may be caused by various factors. It is, for example, likely that children's degree of language experience in the respective languages (e.g., frequency and quality of language input and output, language proficiency) has a significant influence on the rate of acquisition in each language. The longer and the more frequently a child is exposed to and uses a language, the larger the likelihood that they also had more opportunities to stabilise and refine their phonological representations and to practice motor planning and execution skills to produce utterances in this language accurately (Davis & Bedore, 2013; Pearson et al., 1997). This may especially apply to the results of quantitative measures (number of PhonVar and InfVar and PCC-A) in the present research which revealed an overall trend for better performances in German than in Turkish since this was predominantly only significant for older age groups. It could thus be assumed that older children spent proportionally more time in institutional care (i.e., German-speaking environment) than at home which may have increased both their amount of exposure to and usage of German but also the importance of German for their everyday communication settings (Pearson, 2007; Scharff Rethfeld, 2013). This in turn may have enhanced a faster approach to an adult-like pronunciation level in German than in Turkish (see Sections 5.4.2 and 5.4.3 for more detailed discussions).

The different performances in each language may also be a result of a language-specific effect of the phonology in children's ambient languages. The earlier acquisition of the sounds /z, h/ in German than in Turkish, for example, may be explainable by phonological saliency (So & Dodd, 1995; Zhu & Dodd, 2000b). The phoneme /h/ only occurs in the onset in German whereas it can occur in all word-positions in Turkish. Similarly, /z, s/ can occur in all four positions in the word in Turkish but their occurrence in German is distributed over syllable positions with /z/ occurring in the onset and /s/ in the coda only (Fox, 2004). Thus, the phonotactic restrictions for these sounds in German may have increased their saliency and facilitated their acquisition for Turkish-German bilingual children. This finding would thus provide an argument against formalist approaches (e.g., Chomsky & Halle, 1968; Jakobson, 1968) which assume that children have an innate universal knowledge of language which unfolds with experience and leads to an invariable acquisition process of sounds across languages (see Section 1.1). Furthermore, it has clinical implications. Although languages share certain linguistic features it cannot necessarily be expected that these are acquired simultaneously across languages within a bilingual child. SLT's examining a bilingual child need to bear this in mind.

Findings regarding qualitative differences in children's phonological performances (e.g., using phonological patterns in only one language although they could occur in both languages) seem to also provide support for the fact that bilingual children treat their languages differently and are likely to store and process the linguistic information about their two languages separately, i.e., in two phonological systems (Holm & Dodd, 1999c; Paradis, 2001; see Section 1.2.1). However, to pursue the question on the use of one system or two further, more detailed investigations on the individual level would need to be carried out. It would, for example, be revealing to find out whether children substituted shared consonants in the same or different ways across their languages as was explored, for example, by Anderson (2004). For those analyses, however, it would be necessary to take into account children's degree of language experience in the respective languages as this may confound the results (see following section).

5.4 Potential factors influencing children's phonological performances

Results of the single-word naming assessments revealed large standard deviations in Turkish-German bilingual children's phonological performances across measures indicating a great variability in the data. Possible influential factors on monolingual phonological acquisition (e.g., age, gender, language input) have frequently been discussed in the literature (e.g., Dodd et al., 2003; Menn et al., 2013; Vihman, 1996). Due to the heterogeneity of the population further potentially influencing factors (e.g., AoA, language exposure and usage patterns per language, language proficiency per language) came to the fore for bilinguals (e.g., Goldstein et al., 2010; Holm, 1998; Morrow et al., 2014). As reported and discussed in previous sections, children's phonological performances in both languages strongly correlated with their chronological age. Hence, to explore further factors that may have an effect on the variation discovered in Turkish-German bilingual children's phonological productions, gender-related, input- and output-related as well as language proficiency-related differences in children's performances were investigated separately.

In order to gain insight into the influencing factors on the overall degree of variability in children's phonological productions the three holistic measures *number of PhonVar, number of InfVar* and *PCC-A-scores* were used as outcome measures. Whereas performances on the number of InfVar and PCC-A have been discussed in Section 5.1, those on the number of PhonVar have not been explicitly addressed yet. To clarify why they are included here, the measures' benefits over the PCC-A is illustrated and children's performances on the number of PhonVar are briefly summarised. With the number of PhonVar all instances of phonologically variated speech in the children's productions are counted (see Section 3.4.4). Thus, it allows for an analysis of the degree of inaccuracy in children's speech based on the whole word and disengages from the pure segmental perspective as is the case for PCC-A. Similar to the number of InfVar, children demonstrated a very large number of PhonVar in both of their languages and across age groups. The number of PhonVar significantly decreased with age but was still high at the age of 5;0 - 5;5 years (i.e., $M_{German} = 21.9$, $M_{Turkish} = 43.3$), thus suggesting a high degree of phonological inaccuracy until the oldest age group. Further, children's performances significantly differed across their languages and, therefore, may suggest an influence of socio-linguistic variables. Therefore, the number of PhonVar were included in the analyses of influential factors which are discussed in the following.

It should be noted, though, that existing studies focusing on influential factors of children's speech production varied regarding the ways they investigated the effect of variables (i.e., via correlations or regression analyses and by exploring differences between groups) and whether they investigated language input and output aspects separately or combined (e.g., as language experience or language dominance). Comparisons are primarily made with studies exploring differences but where these were not available projects investigating correlations/regressions were considered instead. Consequently, the comparisons in the following need to be considered and interpreted with great caution.

Additionally, it should be borne in mind, that sample size differences across studies existed and that especially in the present research the overall sample size was relatively small. Effects resulting from this (e.g., the absence of a clear influence of a certain variable) were, therefore, plausible.

5.4.1 Influence of gender on children's phonological performance

Based on results of previous research regarding gender-related effects (e.g., Dodd et al., 2003; Ege, 2010; Schäfer & Fox, 2006) it was hypothesised that girls would show a tendency to outperform boys but that effects may be only marginal. In fact, results showed that on average boys performed weaker than girls. This difference, however, was only statistically significant for the number of InfVar in the children's Turkish productions. The effect size was small indicating that gender only explained a small part of the variance in this outcome measure. This result is generally in line with findings in the relevant literature. Regarding different phonological skills, gender-specific differences in monolingual and bilingual children's performances usually favour girls (Ege, 2010) but only rarely reached statistical significance (Dodd et al., 2003; Munro et al., 2005; Schäfer & Fox, 2006). In the case they did reach significance, a gender-related effect was often confounded by language dominance or age effects (Munro et al., 2005). The absence of a clear effect for both languages and on different phonological measures in the present research thus suggests a negligible gender-related difference in Turkish-German bilingual children's phonological skills, which confirms the postulated hypothesis.

5.4.2 Influence of the proportion of language input and output on children's phonological performance

Children of the present research varied greatly regarding the number of hours they were exposed to and used German and Turkish per day. It was hypothesised that children with proportionally more input in German would outperform those with more input in Turkish on German phonological skills and vice versa. Children with an equal amount of input in both languages were expected to outperform children with more input in Turkish on German phonological skills and outperform children with more input in German on Turkish phonological skills. The same predictions were made for effects of the relative amount of language output.

When each of the phonological measures used in German and Turkish (i.e., PhonVar, InfVar and PCC-A-scores) was compared across the different input and output categories (see Section 3.2.3), significant differences were found in the children's German skills only. Proportionally, more input in German resulted in better phonological performances in this language than compared to more input in Turkish. The same held true for differences between language output categories. Since effects were only related to German, the data did not confirm the first general hypothesis (i.e., proportionally more input/output in a certain language results in better phonological skills in this language than receiving more input/output in the other language).

Studies that investigated the effect of language exposure and/or language usage on children's linguistic performance often either found no effect (Goldstein et al., 2010; Goldstein et al., 2005; Kim et al., 2016) or reported effects on both languages of their participants. The latter included significant correlations between language exposure and children's productive vocabulary (Parra et al., 2011; Pearson et al., 1997), significant differences on PPC-scores, the age of suppression of phonological patterns, as well as early and late acquired sounds between groups of children with a different language dominance (Law & So, 2006; Ruiz-Felter et al., 2016), and significant correlations between relative language exposure and phonological memory skills (NWR-performance; Parra et al., 2011). Only a few studies (e.g., Gutiérrez-Clellen & Kreiter, 2003) revealed different effects of language input/output patterns on a bilingual's languages as in the present research. Thus, the question arises why in some cases neither the amount of language input nor the relative amount of language output seem to have an effect on children's linguistic skills – in our case the Turkish language.

One possible explanation for this is that the categories formed did not correspond to natural categories or were not representative of language thresholds for Turkish (Goldstein et al., 2010). Maybe the difference in amount of hours between more input/output in Turkish and equal amount of input/output in both languages needs to be larger to make a difference in the children's Turkish skills. Another possible explanation might be that parents' estimations regarding the amount of German input and output were more realistic than their estimations regarding input and output in Turkish (Goldstein et al., 2010; Goldstein et al., 2005; Gutiérrez-Clellen & Kreiter, 2003). Conceivably, the amount of input and output in German was easily derived from the number of hours the child attended nursery per day, plus (if applicable) some additional time at home. However, parents might have found it difficult to decide how many hours of the time spent at home or with relatives/friends their child heard or spoke Turkish as it was often reported to co-occur with German (cf. Hammer & Rodriguez, 2012). This difficulty would not be surprising since code-mixing and code-switching are common features in bilingual conversations (Grosjean, 2013; Küppers et al., 2015; Romaine, 1995) and were also occasionally observed in children of the present research during the assessment situations (e.g., naming an item in German when asked to name it in Turkish).

Furthermore, considering the amount of input and output separately may not be sufficient to reflect the whole picture of children's language learning environment as the two are intrinsically tied to one another (see e.g., Pearson's input-proficiency-use cycle; Pearson

2007). Language input and output patterns may need to be combined as one variable in order to find significant differential effects (Bedore et al., 2012; Goldstein et al., 2010). Children with more input in German were likely to have had the opportunity to also speak more German (e.g., in the nursery where German is the only language of communication). Children with more input in Turkish, however, might not have automatically spoken more Turkish than German as they may have been attracted more by the language of the environment and thus used German even when addressed in Turkish by their interlocutors. The status of the German language in society is considerably higher than that of a minority language (Scharff Rethfeld, 2013) so it is likely that children may have preferred to use German in bilingual contexts (De Houwer, 2007; Hammer et al., 2011; Pearson, 2007; Romaine, 1995).

To explore this assumption further, correlations between children's amount (i.e., hours) of input and output in both languages were run. Results indicated that the more input children received in German the more output they produced in this language and the less output they produced in Turkish. However, the reverse was not the case since children's amount of input in Turkish and output in German was uncorrelated (see Appendix V for a presentation of these correlations). Thus, there were some children who experienced more input in Turkish but produced more output in German. This would be in line with findings presented earlier in this chapter which showed that children's German skills were on average significantly better than their Turkish phonological skills – irrespective of the applied measure (i.e., consonant and CC inventories, PCC-A-scores, number of PhonVar, number of InfVar and age of suppression of phonological patterns). It might further explain why children with more input and those with more output in Turkish did not significantly differ in their German phonological skills from children with an equal amount of input and/or output in both languages. As indicated by Pearson (2007), the attraction of the environmental language and attitudes towards learning it, may be very powerful.

The absence of significant group differences between the more-input-in-German and equal-input-in-both-languages categories further suggests that the amount of input in German needs to pass a certain threshold to reflect better phonological abilities in German (Gathercole & Hoff, 2007; Pearson, 2007; Pearson et al., 1997). However, besides these considerations it should be borne in mind that children in the three input categories differed significantly regarding age, with children hearing more German being on average the oldest. As age was found to highly correlate with children's phonological performances in both languages (in the present research but also across projects, e.g., Grech & Dodd, 2008; Munro et al., 2005), it might have confounded the effect of input in this research. The applied non-parametric tests, however, did not allow for controlling for age during the analyses of differences.

Concerning within-group differences, the data revealed that equal exposure to both languages did not necessarily lead to balanced phonological skills in Turkish and German. Children who received an equal amount of input in Turkish and German as well as those who used both languages equally frequent demonstrated phonological skills in favour for German on some measures. Comparable findings regarding this imbalance were reported for bilingual Singaporean children’s vocabulary scores (Dixon, Wu, & Daraghmeh, 2012).

Nevertheless, within-group differences suggest that proportionally more input/output in language A leads to on average better phonological abilities in language A (within an individual) compared to language B which points into the same direction as findings from other studies (see, e.g., Law and So (2006) as well as Munro et al. (2005) for differences regarding phonology, Parra et al. (2011) for correlations regarding phonology; and see Scheele, Leseman and Mayo (2010) for relationships between L1 and L2 input patterns and vocabulary skills). However, applying post hoc analyses to their findings Law and So (2006) found that Cantonese phonology generally developed faster than Putonghua ruling out language dominance as the only possible influential factor on children’s phonological skills. They suggested that the complexity of the phonological system (which they admit is difficult to determine) may play a crucial role as well. Similarly, the proportion of language input/output in the present research does not seem to be the only influential factor. When considering the mean performances across languages and categories in the present research, though, it seems as if within-group differences were mainly caused because German skills were influenced by language input and output (and maybe by language status) since children’s Turkish performances were relatively stable across categories. This implies that the amount of language input/output affected the phonological skills in Turkish to a lesser extent so that other influential variables need to be considered as well (e.g., low familiarity with the items in the tests, cf. especially Section 3.3.1.3; effects on specific phonological output rather than holistic measures; language proficiency). An influence by those would reinforce the notion that phonological acquisition is a multi-faceted process in which both child-internal and external factors as well as their interplay during acquisition play an important role as suggested in the emergentist approach (Davis & Bedore, 2013). As outlined at the beginning of this chapter an investigation of the influence of language proficiency was incorporated in this research and is discussed in the following section. Additional potential factors (e.g., the ones outlined above), however, would need to be explored in future research.

5.4.3 Influence of language proficiency on children's phonological performance

There is consensus in the literature that phonological development interacts with lexical and grammatical development (Davis & Bedore, 2013; Roulstone, Loader, Northstone, & Beveridge, 2002; Smith et al., 2006; Stoel-Gammon, 2011; Vihman, 2014). Thus, it seemed important to explore whether children's phonological skills in Turkish and German were influenced by their language proficiency in these languages. It was predicted that language proficiency (as measured by parental report) had a significant effect on children's phonological abilities in the same language. Following Kehoe's (2011) commentary and findings of vocabulary studies (Scheele et al., 2010) it was further hypothesised that due to a positive transfer of linguistic skills – in this case articulatory practice and the development of speech motor skills – children's phonological skills in one language may additionally be predicted by their language proficiency in the other language.

The analyses revealed that children's phonological abilities in German and Turkish were indeed affected by their language proficiency in the same language. In general, the more advanced the children's language ability was the higher their segmental accuracy (i.e., PCC-A-scores) and the lower their number of PhonVar and InfVar in this language. However, not all proficiency categories differed significantly from one another and generally more outcome measures in German than in Turkish were affected by language proficiency.

These findings mainly confirm the first of the above stated hypotheses and are broadly consistent with Goldstein et al.'s (2010) study on Spanish-English bilingual children who found that participants' phonological skills could be predicted by language proficiency as measured by parental report as well as by direct assessment (MLUw). Further, Cooperson et al. (2013) confirmed strong correlations between Spanish-English bilingual children's phonological scores and their morphosyntactic skills as well as low correlations with children's semantic abilities. Similar to the present research, the effects of language proficiency identified in Goldstein et al.'s (2010) study were more distinct on one language (Spanish) than on the other (English). The authors expected this language-specific outcome due to the general structural differences in the phonologies of the two languages. Given the structural differences between Turkish and German phonologies (especially regarding phonotactics) as presented in Section 1.3, these might have had an impact on the language-specific outcome in the present project as well.

Possible explanations for categories not differing significantly might be that either parents found it challenging in general to rate their children's language proficiency accurately (Bedore et al., 2012; Gutiérrez-Clellen & Kreiter, 2003) or the provided categories were not appropriately meeting the actual receptive and expressive competences for every child (Goldstein et al., 2010; Yavaş, 1998). It is imaginable that a direct measure of language proficiency – as, for example, the MLUw – would have supplied a more reliable source for measuring the influence of language proficiency on phonological skills in both languages (Bedore et al., 2012; Goldstein et al., 2010). Including direct measures of language proficiency would thus be recommended for future investigations. However, since mainly children with very good language proficiency differed from those with weaker proficiency, it may also be that language proficiency needs to reach a certain threshold to have an influence on children's performances on phonological tasks (Pearson, 2007). Another aspect that needs further investigation is whether effects would be seen more clearly if language proficiency were measured separately for vocabulary and grammatical skills – rather than combined as in the present research. In this respect, it would also be interesting to see if vocabulary and grammatical skills affected German and Turkish phonological skills differently given the prominent agglutinative nature of the Turkish language.

Considering within-group differences, the data do not only suggest a language-specific effect of language proficiency on the children's phonological skills but also seemed to provide some evidence for cross-linguistic influences. Children who were rated to have high language proficiency in German exhibited weak phonological skills in Turkish and children with satisfactory to poor language proficiency in German demonstrated better phonological skills in Turkish than in German. The effect, however, was more distinct for German than for Turkish proficiency. This finding seems to confirm the second hypothesis (i.e., phonological skills in one language may be predicted by language proficiency in the other language) but presents a different direction than in Scheele et al.'s (2010) work who found that lexical skills in L1 support the acquisition of lexical skills in L2. It is possible though, that a negative cross-linguistic effect of language proficiency may be age-related, in that younger children's development in language A may be enhanced by acquired structures in language B, whereas those structures may act interferingly at a later (more advanced) phonological stage (cf. Scheele et al., 2010). Given the age differences between Scheele et al.'s (2010) study (i.e., 2;11 - 4;1 years) and the present research (3;0 - 5;5 years) this notion seems likely but requires further investigation.

A closer consideration of the children's phonological performances in each of the language proficiency categories, however, discloses that children's performances in Turkish were relatively stable across categories. Therefore, within-group differences in the children's performances in Turkish and German were probably mainly caused by German phonological skills being influenced by language proficiency. Similar to the effect of language input/output this suggests that language proficiency (as measured in this research) is not the only influential factor of children's phonological skills (cf. Law & So, 2006) and that especially children's performances in Turkish seem to be impacted by other variables (see discussion in Section 5.1.2).

Hence, considering the outcomes of the present research as well as those in other studies, it could be concluded that the relationship between phonology and other language domains as found for monolingual children is likely be important during bilingual phonological acquisition as well which would support the interdisciplinary view proposed by functionalists, especially the emergentist approach (Davis & Bedore, 2013). Results further indicate that languages may be affected to a different degree by the individually identified influential factors.

5.5 Summary and conclusions

This chapter has described and discussed the characteristics of the typical phonological acquisition in Turkish-German bilingual children aged 3;0 - 5;5 years. Results confirmed all of the postulated hypotheses and helped to answer the research questions.

With regard to research question 1a, data revealed that Turkish-German bilingual children's rate of phonological acquisition appears to be on average decelerated in both languages if compared to that of monolingual children. However, due to a large inter-individual variability, the fact that the pace of acquisition is commensurate with that reported for other bilinguals, and the observed differences in the rate of acquisition across languages, it was concluded that children's acquisition was not considered to be overall clinically delayed but rather typically-paced for a bilingual population. The large inter-individual variability, however, suggested that children's rate of acquisition was influenced by child-internal and/or -external factors, some of which were then explored further (see further below).

Qualitatively, structural and systemic phonological patterns as well as CLT could be observed in both of the children's languages. Some phonological patterns were known from typically developing monolinguals and pilot studies with Turkish-German bilingual children, others were unknown from previous research. However, not all patterns were frequently occurring (i.e., within and across age groups), which, with regard to research question 1b, impeded the definite assigning of some of the observed patterns to the typical phonological acquisition in Turkish-German bilinguals. Follow-up assessments (see Section 7.2) and ideally larger cohorts were considered to be necessary to investigate the presence and permanence of these patterns further – especially since knowledge of typical phonological behaviour (incl. the knowledge of phonological pattern use) is crucial for the identification of SSD (Dodd, 1995).

Regarding research question 2 data revealed that out of the five examined influential factors on children's phonological acquisition, chronological age, proportion of language input and output as well as language proficiency seemed to have a differential and mainly language-specific effect. Gender, in contrast, only played a marginal role. Given some age differences between some of language input categories, however, further research that controls for an effect of age is necessary to validate findings. Overall, results showed that the factors influencing Turkish-German bilingual children's phonological skills are multi-faceted with none of the factors investigated explaining differences in children's performances across both languages in full. Hence, additional child-internal and environmental variables as well as their weighting in the acquisition process need to be considered and explored in future investigations.

6 RESULTS II – Developmental progress in the phonological acquisition in Turkish-German bilingual children from t1 to t2

In this chapter, the phonological performances at t1 and 12 - 15 months after the initial assessment (t2) are presented for those 43 children who had participated in the follow-up study to describe Turkish-German bilingual children's phonological progress over time. Thus, results presented for t1 in this chapter are not identical to t1-results presented in chapter 4 as here data are only included for the subgroup of children who participated at t1 and t2.

Results in this chapter are reported to specifically address research question 3 (as outlined in Section 2.3) and include the following aspects:

1) Progress in the acquisition of consonants from t1 to t2
2) Progress in the achievement of speech accuracy scores (PCC-A) from t1 to t2
3) Progress in the acquisition of consonant clusters from t1 to t2
4) Development in the observed PhonVar and their age of suppression from t1 to t2.

Each section starts with a short summary of the level of acquisition at t2 before reporting on children's progress over time. Throughout the chapter results for German are presented first and results for Turkish second. However, combined presentations are applied whenever direct comparisons are appropriate. The presentation of data in this chapter is structured by children's age at t2 and the associated age groups. Their respective data/performances from t1 are reported in columns/rows labelled as "t1", however, children's exact age at t1 is not presented. Further, the same scoring criteria and definitions of measures as at t1 (outlined in Section 3.4) were applied to all measures presented in this chapter.

6.1 Progress in the acquisition of consonants from t1 to t2

For the evaluation of children's consonant acquisition at t2, consonant inventories were compiled per participant and analysed separately for German and Turkish.

6.1.1 German consonants

As Table 6.1 shows, children had acquired (75%-criterion) and mastered (90%-criterion) all German consonants except for /ç/ and /v/ which were not mastered by the age of 6;0 - 6;5 years at t2. However, not all the acquired/mastered phonemes in the younger age groups were also acquired/mastered by the older age groups and the data demonstrate no significant correlation between the children's age and the number of acquired German consonants at t2, $r_s = -.064$, $p = .683$.

Table 6.1: Acquisition and mastery of German consonants by children of the t2-cohort at t1 and t2

Age group	4;0 - 4;5		4;6 - 4;11		5;0 - 5;5		5;6 - 5;11		6;0 - 6;5	
N	4		9		11		13		6	
Cons. \ T	**t1**	**t2**	**t1**	**t2**	**t1**	**t2**	**t1**	**t2**	**t1**	**t2**
m	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered
n	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered	mastered
ŋ	acquired	mastered	mastered	mastered	acquired	mastered	acquired	mastered	mastered	mastered
p	mastered	mastered	acquired	acquired	mastered	mastered	mastered	mastered	mastered	mastered
b	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered
t	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered
d	mastered	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered
k	mastered	mastered	acquired	mastered	mastered	mastered	acquired	mastered	mastered	mastered
g	not acquired	mastered	not acquired	acquired	acquired	mastered	not acquired	mastered	mastered	mastered
f	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered
v	not acquired	mastered	not acquired	mastered	acquired	mastered	acquired	acquired	mastered	acquired
s	acquired	mastered	acquired	mastered	acquired	mastered	mastered	mastered	mastered	mastered
z	acquired	mastered	mastered	acquired	mastered	not acquired	mastered	acquired	mastered	mastered
ʃ	not acquired	mastered	not acquired	acquired	not acquired	acquired	not acquired	acquired	acquired	mastered
ç	not acquired	acquired	not acquired	not acquired	not acquired	acquired	not acquired	acquired	not acquired	acquired
x	acquired	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered
ʁ	acquired	mastered	not acquired	acquired	acquired	acquired	acquired	acquired	mastered	mastered
h	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered	mastered
l	mastered	mastered	mastered	mastered	mastered	mastered	acquired	mastered	mastered	mastered
j	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered	mastered
p͜f[1]	not acquired	mastered	not acquired	acquired	not acquired	mastered	acquired	acquired	mastered	mastered
t͜s	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	acquired	mastered

Note: **Cons.**: consonant, **T**: time-point, **shared consonants with Turkish**, consonant mastered by age group, consonant acquired by age group, consonant not acquired by age group

[1]In word-initial position [f] was accepted as phonemically correct for /p͜f/ since this is the variation of German spoken in the North of Germany, where data have been collected.

As at t1 (see Section 4.1.1), consonants that were not acquired by an age group only belonged to the sound classes of fricatives and affricates.

<u>Developmental progress in the acquisition of German consonants from t1 to t2</u>

Most participants (*n* = 28, 65.1%) increased their number of acquired consonants between the initial assessment and its follow up 12 - 15 months later. Despite this, eight of them had at least one sound missing from their inventory that they had already acquired at t1.

For eleven participants (25.6%), representing all age groups, the number of acquired consonants in German remained the same over t1 and t2. Seven of these children had already acquired all consonants at t1. Three of them had acquired the same sounds at t1 and t2, whereas one additional child had indeed the same number of consonants acquired at t1 and t2 but these differed in two sounds.

For four children (9.3%) the number of acquired consonants decreased from t1 to t2. For two of them at least one missing consonant was the same as the missing consonants from t1. The other two children had no sounds missing from their inventories at t1 but two and three, respectively at t2. It has to be noted, though, that in all four cases, at least one consonant was not acquired at t2 because the cut-off was just missed by a count of one.

When considering the whole t2-sample, the overall difference between the number of acquired consonants at t1 (*Mdn* = 19) and t2 (*Mdn* = 21) was statistically significant, Wilcoxon signed-rank test: $z = -4.16$, $p < .001$, $r = -.63$.

6.1.2 Turkish consonants

Turkish-German bilingual children had acquired all Turkish consonants except /z, ʒ/ and had mastered the majority of acquired consonants except for /ʃ, ɾ, h/ at the age of 6;0 - 6;5 years (see Table 6.2). However, as for the German data, not all of the acquired/mastered consonants in the younger age groups were also acquired/mastered by the older age groups and the data do not show a statistically significant correlation between the children's age and the number of acquired Turkish consonants at t2, $r_s = -.187$, $p = 231$. Except for /g/ in the 5;0 - 5;5-year-olds all missing consonants were either fricatives of language-specific consonants.

Table 6.2: Acquisition and mastery of Turkish consonants by children of the t2-cohort at t1 and t2

Age group	**4;0 - 4;5**		**4;6 - 4;11**		**5;0 - 5;5**		**5;6 - 5;11**		**6;0 - 6;5**	
N	**4**		**9**		**11**		**13**		**6**	
Cons. \ T	**t1**	**t2**	**t1**	**t2**	**t1**	**t2**	**t1**	**t2**	**t1**	**t2**
m										
n										
p										
b										
t										
d										
k										
g										
f										
v								*		
s										
z										
ʃ										
ʒ[1]										
h										
ɾ										
l										
j										
t͜ʃ										
d͜ʒ										

Note: **Cons.**: consonant, **T**: time-point, **shared consonants with German**, consonant mastered by age group, consonant acquired by age group, consonant not acquired by age group, *is influenced by missing data, [1]This phoneme only occurs in loan words and according to Ege (2010) and Lewis (1967) is often realised as /d͜ʒ/, even by adult speakers of Turkish.

Developmental progress in the acquisition of Turkish consonants from t1 to t2

The number of acquired Turkish consonants increased from t1 to t2 in 26 participants (60.5%). Six of these children, however, had at least one sound missing from their inventory which they had already acquired at t1. For seven participants (16.3%) the number of acquired consonants remained the same over the two assessment points. In two children, these sounds were identical with the consonants they had acquired at t1. In the other five children, however, at least one consonant was missing from their inventories which they had already acquired at t1. Only one of the children who demonstrated acquisition of the same number of consonants at t1 and t2 showed this for both Turkish and German. The remaining six children only showed this for one of their languages.

Ten participants (23.3%) had acquired fewer Turkish consonants at t2 than at t1. For eight of these children there was an overlap of one to six sounds among the ones which were missing from their Turkish inventories at t1 and t2. The number of additional/newly missing consonants ranged from one to three. For two children, however, the consonants missing at both time points were not identical at all. It has to be noted, though, that in all but one of the ten children, at least one consonant was not acquired at t2 because the cut-off was just missed by a count of one.

Except for one child, all children who had fewer consonants acquired at t2 only showed this decrease in acquisition in one language.

The differences in the children's number of acquired consonants at t1 ($Mdn = 16$) and t2 ($Mdn = 18$) was statistically significant, Wilcoxon signed-rank test: $z = -2.88$, $p = .004$, $r = -.44$.

6.1.3 Shared consonants

All shared consonants between Turkish and German were acquired in German by the age of 5;5 - 5;11 years at t2 (compare sounds marked in red in Table 6.1 and Table 6.2). In Turkish, however, the consonants /v, z/ were still missing from the children's inventories at this age (although /v/ may have been affected by missing data). Except for /v/ in German and /z, ʃ, h/ in Turkish, all shared consonants were mastered by the age of 6;0 - 6;5 years in both languages. Fricatives that were absent from the children's inventories at t2 were subject to different phonotactic restrictions in Turkish and German. All of them are permitted to occur in the syllable onset and coda in Turkish but only in onset positions in German. Participants had acquired these consonants in all overlapping syllable positions but not in the ones that were Turkish-specific. The consonant /z/, for example was acquired by all age groups in the onset of Turkish words but only for the oldest age group in syllable-final positions.

Similar to t1 (see Section 4.1.3) a large variability in the data could be found. There were children who did not miss any shared consonant in their German or Turkish inventories and those who missed up to three in German and up to six shared consonants in Turkish.

<u>Developmental progress in the acquisition of shared consonants from t1 to t2</u>

Different forms of progress from t1 to t2 could be observed in the children's data:

Twelve children (27.9%) had acquired more shared consonants in both of their languages at t2 than at t1. However, the consonants they had acquired at t1 were not necessarily also acquired at t2. In seven cases the children had at least one shared sound acquired at t1 which was absent from their inventories in one language at t2.

In another twelve children (27.9%) the number of acquired shared sounds increased in only one of their languages from t1 to t2, whereas the number of acquired mutual consonants in the other language remained the same. However, those shared consonants were not necessarily identical at both time-points.

For seven children (16.3%) the number of acquired shared consonants remained the same from t1 to t2 in both of their languages. Five of these children (11.6%) showed acquisition of all shared consonants in both languages at t1 and t2. The remaining two children showed acquisition of the same shared consonants at t1 and t2 in one language but not in the other.

Eight children (18.6%) demonstrated acquisition of the same number of shared consonants at t1 and t2 in one of their languages but fewer acquired shared consonants in their other language. These children showed the decrease in Turkish, whereas the number of mutual consonants in German remained the same between t1 and t2.

Two further children (4.7%) showed an increase in the number of acquired mutual consonants in one language and a decrease in the other language. Two additional children exhibited a decrease in both, Turkish and German from t1 to t2.

6.2 Progress in the achievement of speech accuracy scores (PCC-A) from t1 to t2

To measure children's progress in achieving speech accuracy, their PCC-A-scores at t2 were calculated. Figure 6.1 and Figure 6.2 present children's distribution of PCC-A-scores in German and Turkish per age group at t1 and t2. As the figures show, the variability in children's performances decreased at t2 for German but not in all age groups for Turkish. Due to this, a developmental trend in achieving higher PCC-A-scores with age could only be found for German, $r_s = .338$, $p = .027$ but not for Turkish, $r_s = .227$, $p = .143$.

In general, the children's PCC-A-scores at t2 were significantly higher in German (*Mdn* = 91.87) than in Turkish (*Mdn* = 83.71), Wilcoxon signed-rank test: $z = -5.40$, $p < .001$, $r = -.82$. However, no child achieved a PCC-A-score of 100% in either language, except for one child in the oldest age group in German.

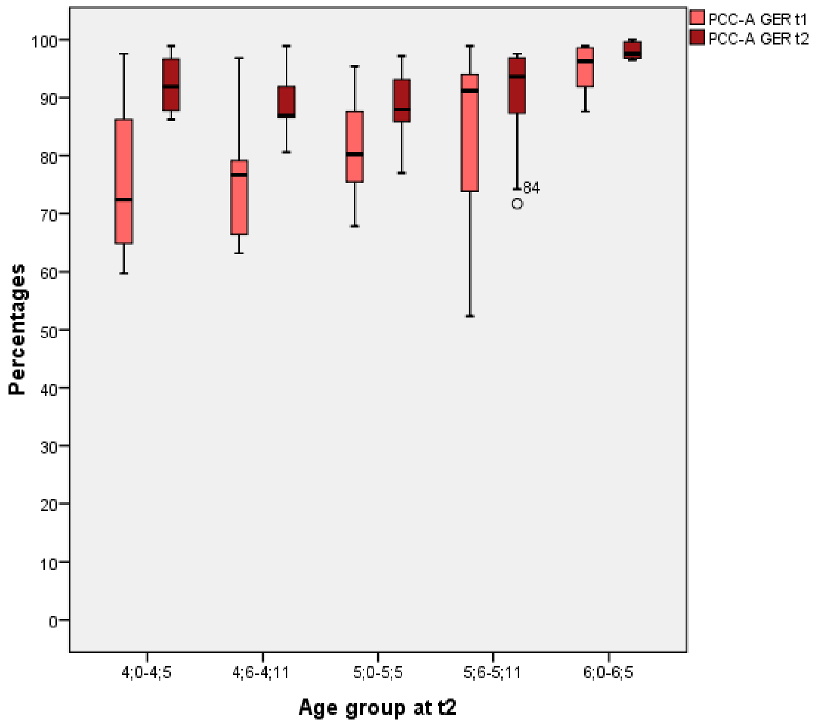


Figure 6.1: PCC-A-scores in German by children of the t2-cohort at t1 and t2

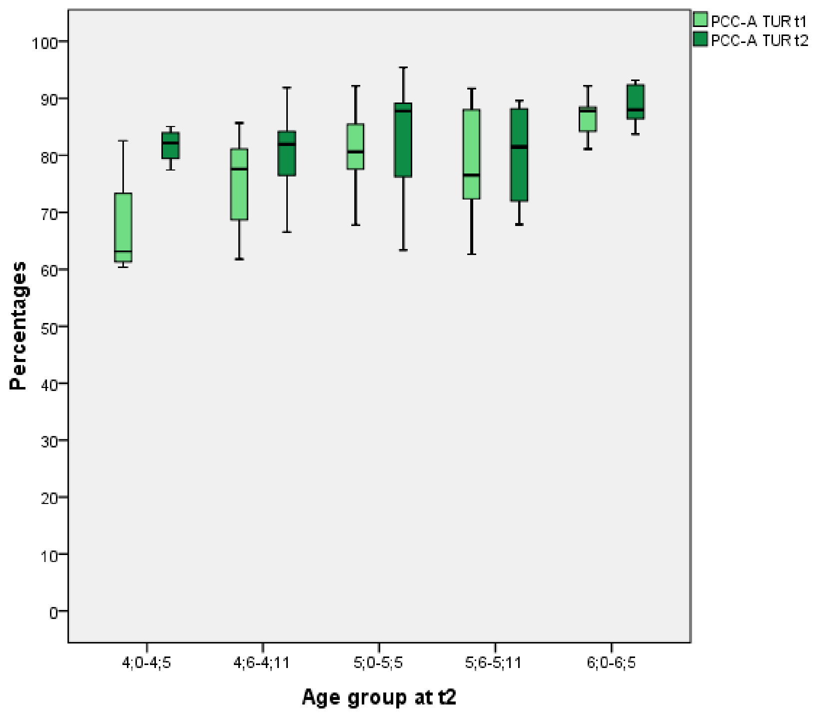


Figure 6.2: PCC-A-scores in Turkish by children of the t2-cohort at t1 and t2

Developmental progress in the PCC-A-scores for German from t1 to t2

Thirty-nine of the children (90.7%) achieved higher PCC-A-scores at t2. The differences ranged from 0.36 percentage points (pp) to 29.70 pp. For four children (9.3%) the PCC-A-scores slightly dropped from t1 to t2. The maximum difference between t1- and t2-scores in these children was 3.89 pp. The overall difference in the children's PCC-A-scores at t1 ($Mdn = 84.45$) and t2 ($Mdn = 91.87$) was significantly different, Wilcoxon signed-rank test: $z = -5.38$, $p < .001$, $r = -.82$.

Developmental progress in the PCC-A-scores for Turkish from t1 to t2

As for German, the majority of children (69%) achieved higher PCC-A-scores in their Turkish productions at t2 than at t1. The differences between t1 and t2 ranged from 0.21 pp to 19.25 pp. For three children, the scores were identical at t1 and t2 and all above 80%. Twenty-three percent of the children ($n = 10$) showed a slightly weaker performance at t2. The maximum difference between t1- and t2-scores in these children was 8.20 pp. The overall difference in the children's PCC-A-scores at t1 ($Mdn = 78.55$) and t2 ($Mdn = 83.71$) was significantly different, Wilcoxon signed-rank test: $z = -3.38$, $p < .001$, $r = -.52$.

All but three children who exhibited higher PCC-A-scores at t2 in Turkish also showed an improvement in their consonantal accuracy in German. The ten children who achieved slightly lower PCC-A-scores in Turkish at t2, also all demonstrated an improvement in PCC-A-scores in German. Two of the children who exhibited lower PCC-A-scores at t2 did so in both of their languages. The remaining children only showed an improvement or decline in one of their languages.

It is generally debateable, however, whether a difference in PCC-A-scores from t1 to t2 of smaller than 2 pp is actually a difference or whether it should be scored as a similar performance since it only meant that 5 out of 283 consonants in German and 4 out of 221 consonants in Turkish have been produced differently from the adult target.

6.3 Progress in the acquisition of consonant clusters from t1 to t2

To assess children's progress in the acquisition of consonant clusters (CCs), onset and coda CC inventories were compiled per participant and analysed separately for German and Turkish.

6.3.1 German onset consonant clusters

As can be seen from Table 6.3, Turkish-German bilingual children had acquired (75%-criterion) all but one German onset CC (i.e., /kv/) by the age of 6;0 - 6;5 years. Except for /kʁ, ʃv/ and the two three-element clusters all CCs were also mastered (90%-criterion) by this age (see criterion 1). Absent CCs mainly contained the phonemes /ʃ/ and /ʁ/ with /kv, t͜sv, p͜fl/ forming the exceptions.

Table 6.3: Acquisition and mastery of German onset CCs by children of the t2-cohort at t1 and t2 – comparison of three different scoring criteria

Age group		4;0 - 4;5		4;6 - 4;11		5;0 - 5;5		5;6 - 5;11		6;0 - 6;5	
CC	*N*	4		9		11		13		6	
	Cr. \ T	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
fl	1										
	2										
	3										
fʁ	1										
	2										
	3										
bl	1										
	2										
	3										
bʁ	1										
	2										
	3										
dʁ	1										
	2										
	3										
gl	1										
	2										
	3										
gʁ	1										
	2										
	3										
tʁ	1										
	2										
	3										
kl	1										
	2										
	3										
kʁ[1]	1										
	2										
	3										
kn	1										
	2										
	3										
kv	1										

Age group		4;0 - 4;5		4;6 - 4;11		5;0 - 5;5		5;6 - 5;11		6;0 - 6;5	
CC	*N*	4		9		11		13		6	
	Cr. \ T	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
	2										
	3										
ʃl	1										
	2										
	3										
ʃʁ	1										
	2										
	3										
ʃm	1										
	2										
	3										
ʃn	1										
	2										
	3										
ʃv	1										
	2										
	3										
ʃp	1										
	2										
	3										
ʃt	1										
	2										
	3										
ʃpʁ	1										
	2										
	3										
ʃtʁ	1										
	2										
	3										
p͡fl[2]	1										
	2										
	3										
t͡sv[3]	1										
	2										
	3										

Note: **Cr.**: criterion, **T**: time-point, **criterion 1**: CC considered to be correct if it is produced phonemically correct, **criterion 2**: CC is considered to be correct if the cluster elements are produced correctly but split up by an epenthetic vowel/Schwa, **criterion 3**: CC is considered to be correct if the number of cluster elements is produced correctly, CC mastered by age group, CC acquired by age group, CC not acquired by age group

[1]/kʁ/ is the only CC that is assessed in a weak syllable (i.e., /kʁokoˈdiːl/). This may have affected the accuracy of production for this cluster and its absence from the children's inventories across all age groups is likely to be an artefact of the test item rather than an acquisition issue.

[2]In word-initial position /p͡f/ is usually deaffricated to /f/ in Northern German. Thus, this realisation was also accepted as phonemically correct (see Sections 1.3.1.4 and 3.4.1).

[3]/t͡sv/ only occurs in a very limited number of child appropriate words.

In most cases children continued to have the underlying representation of CCs acquired before they could produce them phonemically correct. Especially the syllable shape of

CCs consisting of stop/fricative + liquid was earlier acquired than their phonemically correct pronunciation. The difference in the number of acquired CCs with criterion 1 (*Mdn* = 18) and 3 (*Mdn* = 21) was statistically significant, Wilcoxon signed-rank test: $z = -5.34$, $p < .001$, $r = -.81$.

The epenthesis of a vowel/schwa to split up CCs was not frequently present in the children's data at t2. But if the pattern occurred, mainly stop + liquid clusters were affected. A comparison of the performances according to criterion 1 (*Mdn* = 18) with those according to criterion 2 (*Mdn* = 19) also revealed a statistically significant difference, Wilcoxon signed-rank test: $z = -2.43$, $p = .015$, $r = -.88$.

Most of the two-element /ʃ/-CCs were acquired according to the following developmental pattern: acquisition of cluster structure > mastery of cluster structure > phonemic acquisition. However, no significant correlation between children's age and their acquisition of German onset CCs could be found (criterion 1: $r_s = .255$, $p = .099$, criterion 2: $r_s = .246$, $p = .111$, criterion 3: $r_s = .221$, $p = .155$).

Developmental progress in the acquisition of German onset CCs from t1 to t2

When considering all three scoring criteria most children had acquired more CCs at t2 than at t1 (i.e., criterion 1: $n = 31$, 75%; criterion 2: $n = 35$, 81.4%; criterion 3: $n = 37$, 86%). However, there was a small subset of children who showed an identical number of acquired CCs at t1 and t2 (i.e., criterion1: $n = 7$, 16.3%; criterion 2: $n = 6$, 14%; criterion 3: $n = 3$, 7%), although the acquired CCs were not necessarily the same at both time-points. In general, these children had already acquired many CCs at t1.

Additionally, there were a few children who exhibited a slightly smaller number of acquired CCs at t2 than at t1 (i.e., criterion 1: $n = 1$, 2.3%; criterion 2: $n = 2$, 4.7%; criterion 3: $n = 3$, 7%). Most of these children had already acquired a large number of CCs at t1 and the difference in the number of acquired CCs was one or two.

In general, the differences in the acquisition of German onset CCs between t1 and t2 were statistically significant for all three scoring criteria, Wilcoxon signed-rank test:

- Criterion 1: ($Mdn_{t1} = 12$, $Mdn_{t2} = 18$), $z = -5.05$, $p < .001$, $r = -.77$
- Criterion 2: ($Mdn_{t1} = 14$, $Mdn_{t2} = 19$), $z = -4.97$, $p < .001$, $r = -.76$
- Criterion 3: ($Mdn_{t1} = 17$, $Mdn_{t2} = 21$), $z = -5.30$, $p < .001$, $r = -.81$.

6.3.2 German coda consonant clusters

Nearly all assessed German coda CCs were acquired (criterion 1) by the age of 6;0-6;5 years and more than half of them were also mastered at this age at t2 (see Table 6.4). Since there was only one instance of vowel epenthesis in coda clusters, data with this criterion are not presented here. As for the onset CCs, syllable structures of coda CCs were usually acquired before the correct pronunciation of the cluster elements.

Table 6.4: Acquisition and mastery of German coda CCs by children of the t2-cohort at t1 and t2 – comparison of two different scoring criteria

Age Group		4;0 - 4;5		4;6 - 4;11		5;0 - 5;5		5;6 - 5;11		6;0 - 6;5	
CC	*N*	4		9		11		13		6	
	Cr. \ T	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
lç	1	not acquired	not acquired	not acquired	acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	acquired	acquired	acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired
lp	1	mastered	mastered	not acquired	mastered	acquired	mastered	mastered	acquired	mastered	mastered
	3	mastered	mastered	not acquired	mastered	mastered	mastered	mastered	mastered	mastered	mastered
lt	1	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered	mastered
	3	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered	mastered
ns	1	acquired	mastered	not acquired	mastered	not acquired	not acquired	mastered	mastered	mastered	mastered
	3	mastered	mastered	not acquired	mastered	acquired	mastered	mastered	mastered	mastered	mastered
nt	1	mastered	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered
	3	mastered	mastered	mastered	acquired	mastered	mastered	mastered	mastered	mastered	mastered
ŋk	1	mastered	mastered	mastered	acquired	mastered	acquired	acquired	mastered	mastered	mastered
	3	mastered	mastered	mastered	acquired	mastered	mastered	acquired	mastered	mastered	mastered
st	1	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	acquired	not acquired	mastered	acquired
	3	not acquired	not acquired	acquired	not acquired	not acquired	acquired	acquired	not acquired	mastered	mastered
çt	1	not acquired	acquired	not acquired	acquired	not acquired	acquired	not acquired	not acquired	not acquired	acquired
	3	acquired	acquired	not acquired	acquired	not acquired	mastered	not acquired	acquired	acquired	mastered
nst	1	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	acquired
	3	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	acquired

Note: **Cr.**: criterion, **T**: time-point, **criterion 1**: CC is considered to be correct if it is produced phonemically correct, **criterion 3**: CC is considered to be correct if the number of cluster elements is produced correctly, CC mastered by age group, CC acquired by age group, CC not acquired by age group

In most cases where the /lç/-CC was not acquired it was either reduced to one element or realised as /lt͜ʃ/. The latter could reflect a more emphasised articulation of the transition from /l/ to /ç/ with the fricative being coronalised to /ʃ/. Another CC that seemed to be difficult to acquire is the three-element-CC /nst/. Only children in the oldest age group demonstrated acquisition of this CC and its underlying representation of the syllable structure at t2.

Overall, the acquisition of German coda CCs was not found to be following a developmental pattern at t2 (criterion 1: $r_s = -.054$, $p = .730$; criterion 3: $r_s = .072$, $p = .647$) as can be expected from the information in Table 6.4.

Developmental progress in the acquisition of German coda CCs from t1 to t2

According to the results for both scoring criteria most children had acquired more coda CCs or underlying syllable structures of coda CCs at t2 than at t1 (criterion 1: $n = 25$, 58.1%; criterion 3: $n = 28$, 65.1%). A slightly smaller number of children exhibited an equal number of clusters/syllable structures for coda CCs in their inventories as acquired at t2 (criterion 1: $n = 12$, 27.9%; criterion 3: $n = 10$, 23.3%). Only a few children had fewer CC/syllable structures of CCs acquired than at t1 (criterion 1: $n = 6$, 14%; criterion 3: $n = 5$, 11.6%). It must be borne in mind, though, that for each cluster only the first occurrence in the test was evaluated.

The differences in the children's performances at t1 and t2 were statistically significant for both scoring methods, Wilcoxon signed-rank test: criterion 1: ($Mdn_{t1} = 6$, $Mdn_{t2} = 8$), $z = -3.40$, $p < .001$, $r = -.52$ with overall more coda CCs being acquired at t2, criterion 3: ($Mdn_{t1} = 7$, $Mdn_{t2} = 8$), $z = -3.37$, $p < .001$, $r = -.51$ with more underlying syllable shapes for German coda CCs acquired at t2.

6.3.3 Turkish coda consonant clusters

At t2 the children had acquired three out of six assessed Turkish CCs (see Table 6.5). For two further clusters the underlying syllable structure was acquired for at least half of the age groups. Since there was only one instance of vowel epenthesis in coda clusters, data with this criterion are not presented here. The cluster /ft/ was not acquired by any age group.

Table 6.5: Acquisition and mastery of Turkish coda CCs by children of the t2-cohort at t1 and t2 – comparison of two different scoring criteria

Age group		4;0 - 4;5		4;6 - 4;11		5;0 - 5;5		5;6 - 5;11		6;0 - 6;5	
CC	*N*	4		9		11		13		6	
	Cr. \ T	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
lp	1	mastered	acquired	not acquired	acquired	mastered	not acquired	acquired	mastered	mastered	mastered
	3	mastered	mastered	acquired	mastered	mastered	not acquired	mastered	mastered	mastered	mastered
ɾt	1	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	mastered	not acquired	not acquired	acquired	acquired	not acquired	acquired	not acquired	not acquired
ɾk	1	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	acquired
	3	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	mastered
nt͜ʃ	1	acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired
	3	mastered	not acquired	acquired	acquired	acquired	mastered	mastered	not acquired	acquired	acquired
nk	1	mastered	mastered	acquired	mastered	mastered	mastered	mastered	acquired	mastered	acquired
	3	mastered	mastered	acquired	mastered	mastered	mastered	mastered	acquired	mastered	acquired
ft	1	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired
	3	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired	not acquired

Note: **Cr.**: criterion, **T**: time-point, **criterion 1**: CC considered to be correct if it is produced phonemically correct, **criterion 3**: CC is considered to be correct if the number of cluster elements is produced correctly, CC mastered by age group, CC acquired by age group, CC not acquired by age group

Like the acquisition of German coda clusters there was also no statistically significant correlation between the children's age at t2 and the number of acquired CCs in Turkish, $r_s = .232$, $p = .135$ (criterion 1) nor with the number of acquired syllable structures for the assessed CCs, $r_s = .212$, $p = .172$ (criterion 3).

The most prominent observation from this table is that the children's acquisition and mastery of Turkish coda clusters was not stable. In many instances the children had acquired a certain coda cluster at t1 which they were not able to produce at t2.

<u>Developmental progress in the acquisition of Turkish coda CCs from t1 to t2</u>

The number of missing CCs in Turkish did not differ significantly from t1 to t2 for both scoring criteria, Wilcoxon signed-rank test: criterion 1: ($Mdn_{t1} = 4$, $Mdn_{t2} = 4$), $z = -.88$, $p = .405$, $r = -.13$; criterion 3: ($Mdn_{t1} = 4$, $Mdn_{t2} = 4$), $z = -.11$, $p = .916$, $r = -.02$. The numbers of children who had more, less or an equal number of Turkish clusters missing at t1 and t2 were nearly identical (i.e., 16, 12 and 15, respectively).

In case children had an equal number of CCs missing at both time-points the clusters were not necessarily the same.

6.4 Progress in the production of PhonVar from adult-like speech from t1 to t2

For the evaluation of children's progress in producing phonologically variated speech, their number of produced PhonVar, number of InfVar, and (number of different) types of phonological patterns and the ages of their suppression were analysed.

6.4.1 Number of produced PhonVar

The distribution of children's number of PhonVar in German and Turkish at t1 and t2 are displayed in Figure 6.3 (German) and Figure 6.4 (Turkish). These excluded variations resulting from CLT. For both languages a large variability across age groups could be found and only the German data significantly correlated with the children's age at t2: $r_s = -.327$, $p = .033$, Turkish: $r_s = -.193$, $p = .216$. As for t1, the children produced significantly fewer PhonVar in German ($Mdn = 22$) than in Turkish ($Mdn = 38$), Wilcoxon signed-rank test: $z = -3.85$, $p < .001$, $r = -.59$.

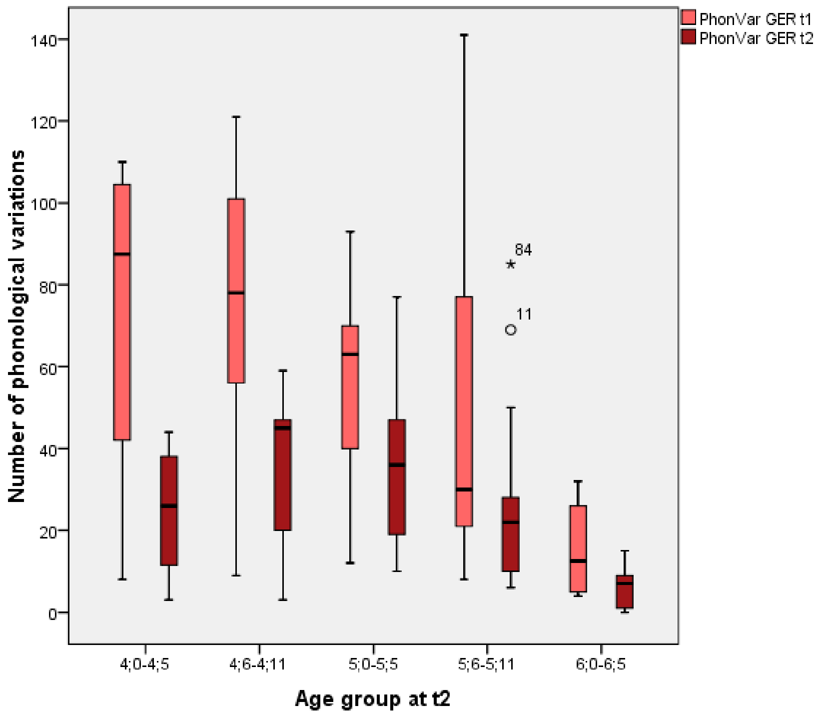


Figure 6.3: Distribution of the number of PhonVar produced in German by children of the t2-cohort at t1 and t2

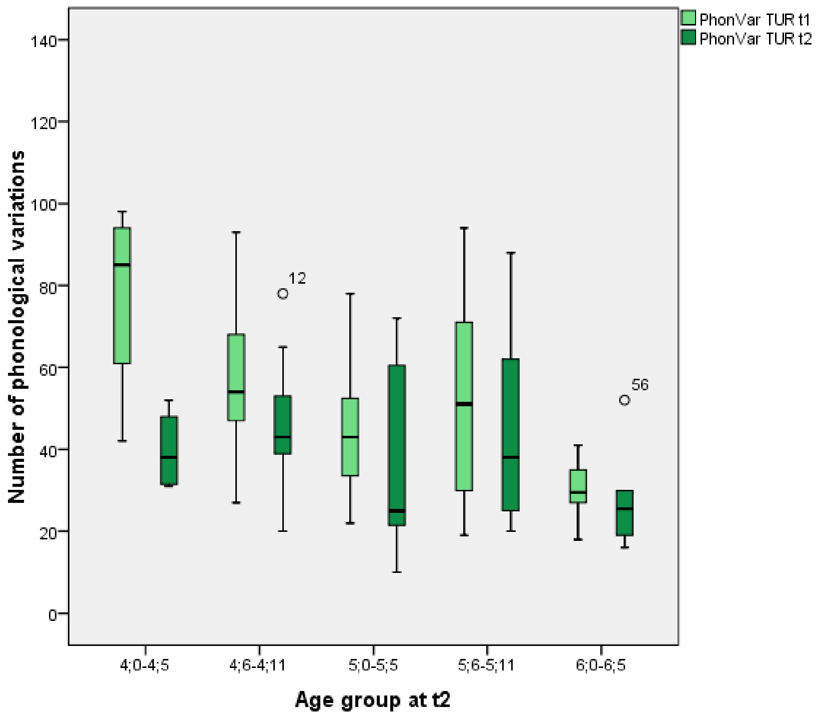


Figure 6.4: Distribution of the number of PhonVar produced in Turkish by children of the t2-cohort at t1 and t2

Developmental progress in the production of PhonVar from t1 to t2

The difference in PhonVar at t1 and t2 had to be larger than 5% to be scored as a difference, otherwise it was considered to be similar. This was applied as there was a disagreement of 5% in the inter-rater reliability for transcriptions.

Across all age groups a decline in the mean number of PhonVar from t1 to t2 could be observed for both languages. The largest decline over time was evident in the youngest age group. Only a small subset of children produced a similar amount or slightly more PhonVar at t2 than at t1 (German: $n = 2$, Turkish: $n = 11$). Despite a large inter-participant variability, the improvement in the children's speech skills from t1 to t2 was highly statistically significant for both languages, Wilcoxon signed-rank test: German: ($Mdn_{t1} = 51$, $Mdn_{t2} = 22$), $z = -5.62$, $p < .001$, $r = -.86$; Turkish ($Mdn_{t1} = 48$, $Mdn_{t2} = 38$), $z = -3.78$, $p < .001$, $r = -.58$.

6.4.2 Number of produced InfVar

On average, InfVar constituted 80.55% (German) and 77.24% (Turkish) of the PhonVar in the sample. The distribution of children's number of InfVar in German and Turkish at t1 and t2 are displayed in Figure 6.5 (German) and Figure 6.6 (Turkish).

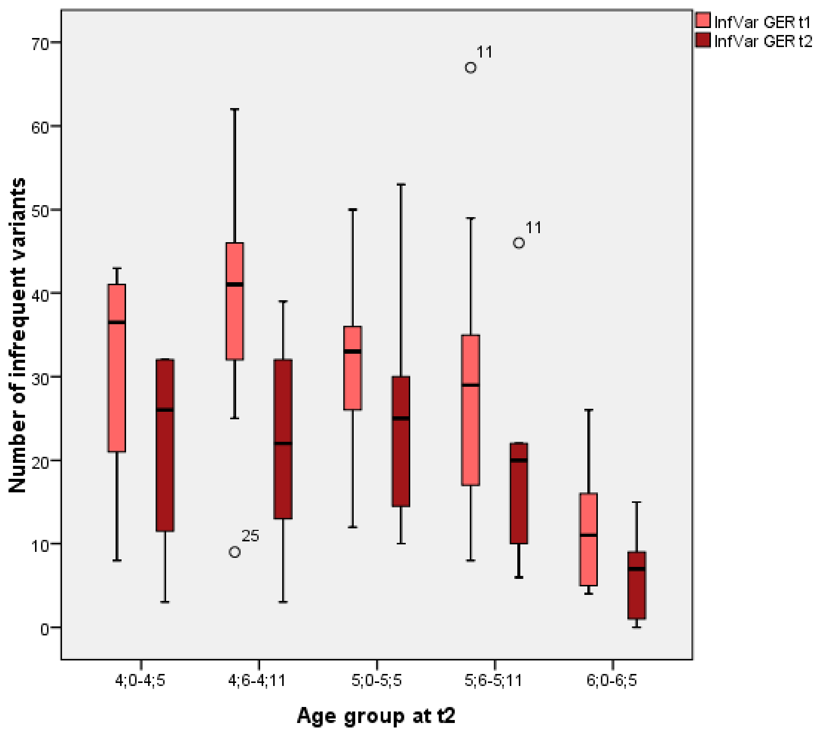


Figure 6.5: Distribution of the number of InfVar produced in German by children of the t2-cohort at t1 and t2

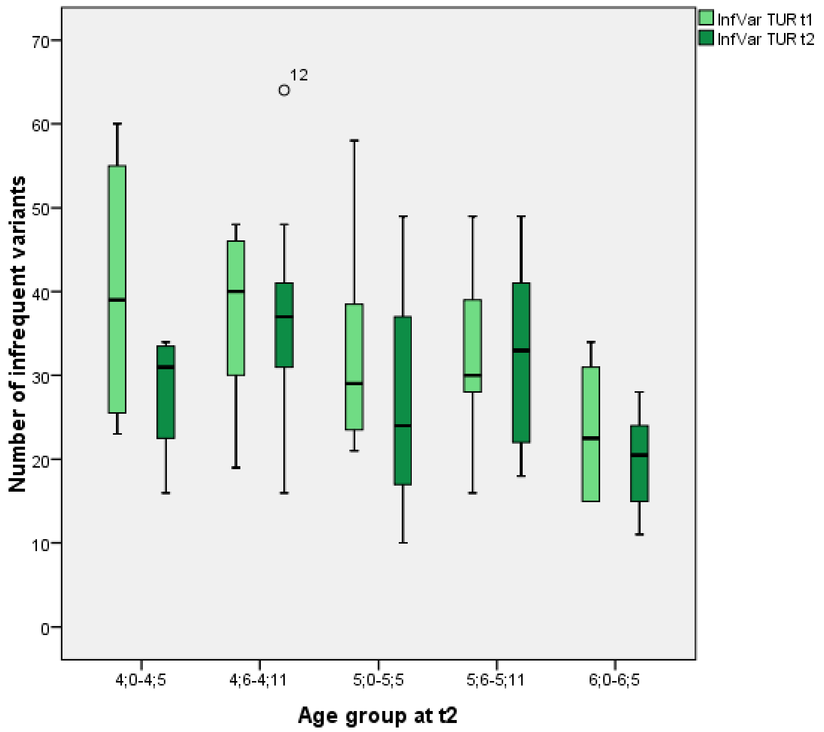


Figure 6.6: Distribution of the number of InfVar produced in Turkish by children of the t2-cohort at t1 and t2

The data show a large variability in the number of produced InfVar across age groups and only the German data correlated significantly with children's age at t2, $r_s = -.407$, $p = .007$, Turkish: $r_s = -.253$, $p = .102$.

Like at t1, the number of InfVar was significantly higher for Turkish (*Mdn* = 28) than for German (*Mdn* = 20), Wilcoxon signed-rank test: $z = -4.17$, $p < .001$, $r = -.64$.

Developmental progress in the production of InfVar from t1 to t2

For both languages, the number of InfVar significantly decreased from t1 to t2, Wilcoxon signed-rank test: German ($Mdn_{t1} = 30$, $Mdn_{t2} = 20$), $z = -4.95$, $p < .001$, $r = -.75$; Turkish ($Mdn_{t1} = 30$, $Mdn_{t2} = 28$), $z = -2.02$, $p = .044$, $r = -.31$. However, the decrease varied considerably from one child to another. Some children reduced the overall number of PhonVar by (more than) half, others reduced them only slightly (Range in German: 1 - 48 and Turkish: 1 - 31).

6.4.3 Production of phonological patterns at t2

This section covers the average number of phonological pattern types in German and Turkish per age group, the individual patterns and their ages of suppression as well as the types of phonological patterns which could be attributed to an interaction between Turkish and German.

6.4.3.1 Number of different pattern types

The proportion of phonological patterns among the children's variations from adult-like speech was 18.3% (German) and 22.7% (Turkish) on average at t2. Table 6.6 presents the number of different phonological pattern types produced per age group by the t2-cohort at t1 and at t2. Instances of CLT were not included in this analysis.

Table 6.6: Average number of phonological pattern types in German and Turkish observed in children of the t2-cohort at t1 and t2

Age group	*N*	Time	Pattern types German				Pattern types Turkish			
			M	***SD***	**Min**	**Max**	***M***	***SD***	**Min**	**Max**
4;0 - 4;5	**4**	t1	4.75	2.95	0	8	4.50	1.80	2	7
		t2	0.25	0.43	0	1	1.50	1.12	1	3
4;6 - 4;11	**9**	t1	4.33	2.31	0	8	2.78	1.62	1	6
		t2	1.56	1.50	0	4	1.33	0.67	0	2
5;0 - 5;5	**11**	t1	2.64	1.82	0	6	1.36	0.77	0	2
		t2	1.09	1.31	0	4	1.18	0.83	0	2
5;6 - 5;11	**13**	t1	1.77	2.39	0	8	1.85	1.61	0	5
		t2	0.92	1.38	0	4	1.46	1.60	0	5
6;0 - 6;5	**6**	t1	0.17	0.37	0	1	0.83	1.07	0	3
		t2	0.00	0.00	0	0	1.33	1.25	0	3

The largest group of children ($n = 20$, 46.5%) produced more phonological pattern types in Turkish than in German at t2. For nine children (20.9%) the opposite was the case. Fourteen children exhibited a similar number of pattern types in German and Turkish. The difference between produced pattern types in German (*Mdn* = 0) and in Turkish (*Mdn* = 1), however, was only marginally statistically significant, Wilcoxon signed-rank test: $z = -1.92$, $p = .057$, $r = -.29$. There was also no statistically significant correlation between the children's age at t2 and the number of produced pattern types in both languages (German: $r_s = -.218$, $p = .160$; Turkish: $r_s = -.002$, $p = .988$).

<u>Developmental progress in the production of phonological pattern types from t1 to t2</u>

On average, children produced significantly fewer pattern types at t2 than at t1 in German, Wilcoxon signed-rank test: ($Mdn_{t1} = 2$, $Mdn_{t2} = 0$), $z = -4.36$, $p < .001$, $r = -.66$ and Turkish: ($Mdn_{t1} = 2$, $Mdn_{t2} = 1$), $z = -2.82$, $p = .005$, $r = -.43$. However, for several children the number of pattern types remained the same at the two time-points (German: $n = 16$, 37.2%; Turkish: $n = 22$, 51.2%). Most of these children exhibited no phonological patterns at t1 and t2. Others continued to show the same patterns as at t1 or, in a few cases, exhibited different pattern types than at t1. Additionally, a small subgroup of participants (German: $n = 2$, 4.7%; Turkish: $n = 3$, 7%) produced more phonological pattern types at t2. However, these 'new' pattern types had also been present at t1 but occurred to a lower frequency (i.e., one to four times).

6.4.3.2 Phonological pattern types and their age of overcome

In this subsection, all phonological pattern types which could not be directly ascribed to an interaction between German and Turkish are presented. The observed phonological pattern types were divided into structural and systemic simplifications.

<u>German</u>

Table 6.7 displays the phonological patterns in German produced by the children at t2 as well as those produced by the same cohort at t1. Structural simplifications were suppressed the latest by 5;6 years. The most prominent pattern among these simplifications was initial CC reduction; however, this pattern only occurred infrequently with a mean of 8.25 instances per child.

Most systemic simplifications were only present in a very small subset of children at t2. They were usually overcome at the latest by the age of 5;6 years, with deaffrication being an exception. Fronting of /ʃ/ was the most frequent systemic simplification in the children's speech at t2 and occurred consistently in most children. Similarly, deaffrication was used consistently to simplify /t͜s/, whereas /p͜f/ was only rarely affected. None of the children in the oldest age group produced any phonological pattern in German.

Table 6.7: Phonological patterns produced in German by children of the t2-cohort at t1 and t2

Pattern	Age / Time	4;0 - 4;5 *n* (%)	4;6 - 4;11 *n* (%)	5;0 - 5;5 *n* (%)	5;6 - 5;11 *n* (%)	6;0 - 6;5 *n* (%)
Structural simplifications						
Initial CC reduction[1]	**t1**	3 (75)	7 (77)	8 (73)	5 (39)	
	t2		4 (44)	3 (27)		
Final CC reduction[2]	**t1**	1 (25)	2 (22)	2 (18)		
	t2		1 (11)			
Weak syllable deletion	**t1**	1 (25)	5 (55)	3 (27)		
	t2					
Intrusive consonant	**t1**	2 (50)	1 (11)	2 (18)		
	t2					
SI consonant deletion	**t1**					
	t2		1 (11)			
Systemic simplifications						
Fronting /ʃ/	**t1**	3 (75)	5 (55)	4 (36)	3 (23)	1 (17)
	t2	1 (25)	2 (22)	3 (27)		
Fronting /k, g, ŋ/	**t1**	2 (50)			2 (15)	
	t2					
Assimilation	**t1**	2 (50)	3 (33)		2 (15)	
	t2					
Fronting /ç/	**t1**		1 (11)		2 (15)	
	t2		1 (11)			
Allophonic use of sibilants	**t1**			2 (18)		
	t2			2 (18)		
Deaffrication /t͜s, p͜f/	**t1**	1 (25)	4 (44)			
	t2		2 (22)		4 (31)	
Devoicing[3]	**t1**	1 (25)	1 (11)			
	t2		2 (22)			
Stopping Fric. & Affr.	**t1**	2 (50)	1 (11)			
	t2					
Metathesis	**t1**		1 (11)			
	t2					
Backing /t, d, n/	**t1**		1 (11)			
	t2					
Affrication	**t1**					
	t2		1 (11)			

Note: **Time**: time-point, **[1]**includes syllable-initial consonant clusters, **[2]**includes syllable-final consonant clusters, **SI**: syllable-initial, **[3]**(de-)voicing also occurs in a range of regional varieties of German (Fox-Boyer, 2014c; Fox, 2007) although not in the ones where the data were collected, **Fric.**: fricatives, **Affr.**: affricates, pattern produced by 10 - 14% of an age group (shown in grey), blank cells: pattern occurred in 0 - 9.9% of an age group.
The table includes a few patterns in the t1-rows that were not reported for the whole t1-cohort in Section 4.3.3.2. This was due to the smaller number of children in the age groups at t2 which allowed the patterns to meet the cut-off criteria.

Developmental progress in the use of phonological patterns in German from t1 to t2

The majority of children (*n* = 25, 58.1%) did not produce any phonological patterns in German at t2. These children belonged to all age groups. In 22 of these children the number of InfVar reduced from t1 to t2 as well. Two children exhibited the same number of InfVar at t1 and t2 and for one child a slightly higher number of InfVar at t2 could be noted.

Overall, the number of children that produced a certain pattern decreased from t1 to t2. Devoicing, deaffrication and the allophonic use of sibilants formed exceptions. Ten children (23.3%) only exhibited phonological patterns that they had already shown at t1. In six of these cases the token of these patterns was lower at t2 compared to t1. For four children, the difference in the tokens was minor (i.e., one to three occurrences more or less than at t1). Further, apart from the allophonic use of sibilants all other phonological patterns that were considered to be deviant for monolingual children (i.e., labialisation, intrusive consonant, metathesis, backing /t, d, n/; see Section 4.3.3.2) decreased over time and were not observable at the group level any more. Instead two patterns emerged in more than 10% of an age group that were not present at the age group level at t1 before (i.e., syllable-initial consonant deletion, affrication).

Turkish

Only two simplification patterns that affected the word structure were present in the children’s Turkish productions at t2: syllable-final consonant deletion and intrusive consonants (see Table 6.8). Both patterns occurred inconsistently in the children’s speech (i.e., only to a maximum of six and seven occurrences, respectively).

The most prominent systemic simplification pattern at t2 was devoicing. It was present in all age groups and occurred to a maximum frequency of twelve instances per participant. Fronting of postalveolar sibilants/affricates and assimilation were also among the most common phonological patterns at t2. Whereas assimilations occurred inconsistently, fronting of /ʃ, ʒ, t͜ʃ, d͜ʒ/ often affected all items with these sounds.

Table 6.8: Phonological patterns produced in Turkish by children of the t2-cohort at t1 and t2

Pattern	Age / Time	4;0 - 4;5 *n* (%)	4;6 - 4;11 *n* (%)	5;0 - 5;5 *n* (%)	5;6 - 5;11 *n* (%)	6;0 - 6;5 *n* (%)
Structural simplifications						
SF consonant deletion	**t1**	1 (25)	4 (44)		3 (23)	
	t2		4 (44)			
Intrusive consonant	**t1**	1 (25)				
	t2		1 (11)			
Systemic simplifications						
Fronting /ʃ, ʒ, t͜ʃ, d͜ʒ/	**t1**	3 (75)	6 (66)	3 (27)	4 (31)	2 (33)
	t2	1 (25)	3 (33)	2 (18)		1 (17)
Liquid deviation /r/ → [l]	**t1**	2 (50)	3 (33)	2 (18)	2 (15)	1 (17)
	t2	1 (25)				1 (17)
Devoicing[1]	**t1**	2 (50)	4 (44)	2 (18)	2 (15)	1 (17)
	t2	2 (50)	3 (33)	6 (55)	6 (46)	2 (33)
Word-final devoicing	**t1**					1 (17)
	t2		1 (11)			1 (17)
Fronting /k, g/	**t1**		1 (11)		2 (15)	
	t2					
Assimilation	**t1**	3 (75)	1 (11)	3 (27)	4 (31)	
	t2	2 (50)	1 (11)	2 (18)	2 (15)	1 (17)
Deaffrication /t͜ʃ, d͜ʒ/	**t1**	1 (25)	1 (11)		2 (15)	
	t2		1 (11)		4 (31)	
Metathesis	**t1**	1 (25)	2 (22)			
	t2				2 (15)	
Gliding /r/ → [j, w]	**t1**	1 (25)	1 (11)			
	t2					
Stopping Fric. & Affr.	**t1**	1 (25)				
	t2					
Backing /s, z/	**t1**	1 (25)				
	t2					

Note: **Time**: time-point, **SF**: syllable-final, **[1]**excluding devoicing of word-intial alveolar and velar stops as these could be dialectal/accentual variations (Brendemoen, 1998), **Fric.**: fricatives, **Affr.**: affricates, produced by 10 - 14% of an age group, blank cells: pattern occurred in 0 - 9.9% of an age group
The table includes a few patterns in the t1-rows that were not reported for the whole t1-cohort in Section 4.3.3.2. This was due to the smaller number of children in the age groups at t2 which allowed the patterns to meet the cut-off criteria.

Developmental progress in the use of phonological patterns in Turkish from t1 to t2

For most phonological patterns, a smaller number of children was found to produce these patterns at t2. Exceptions are general devoicing, deaffrication and word-final devoicing. In total, four patterns were completely overcome by the children at t2. Many children (*n* = 16, 37.2%) exhibited at least one new phonological pattern at t2 in Turkish. Some of them might be explainable by a transfer from German phonotactic constraints to Turkish (e.g., an increase in word-final devoicing). Others did occur at t1 as well but were less frequent than the used cut-off value of five. Fourteen participants (32.6%) only produced phonological patterns that they had already shown at t1 and 13 participants (30.2%) did not show any phonological patterns in Turkish at t2 at all. Further, one of the phonological patterns that were considered to be unknown for typical and atypical development in Turkish at t1 (i.e., initial consonant deletion) could not be found in at least 15% of any age group at t2. The other pattern (i.e., intrusive consonant) was resolved by one child (age group 4;0 - 4;5 years) but produced by another child (age group 4;6 - 4;11 years).

6.4.3.3 Cross-linguistic transfer at t2

To demonstrate the variety of interactions between German and Turkish found to be present in Turkish-German bilingual children's speech, all observed phonetic and phonological CLT instances were reported, even if these were present in less than ten percent of an age group.

German

Overall, children produced fewer types of CLT in their German speech at t2 than at t1 (see Table 6.9) and no new types of CLT occurred. Some patterns ascribable to CLT were overcome by t2 (e.g., schwa/vowel epenthesis in CCs) or were observable in a smaller sample (e.g., /ʁ/ → [ɾ]). Others occurred in slightly more participants compared to t1 (e.g., /ç/ → [ʃ]).

Table 6.9: Cross-linguistic transfer observed in German by children of the t2-cohort at t1 and t2

Pattern	**Age / Time**	**4;0 - 4;5** *n* (%)	**4;6 - 4;11** *n* (%)	**5;0 - 5;5** *n* (%)	**5;6 - 5;11** *n* (%)	**6;0 - 6;5** *n* (%)
Structural simplifications						
Vowel epenthesis in CCs	**t1**	1 (25)	3 (33)	3 (27)	1 (8)	
	t2					
BPR of /ŋ/→[ŋg]*	**t1**				1 (8)	
	t2			1 (9)		
Systemic simplifications						
/ç/ → [ʃ]*[1]	**t1**				1 (8)	1 (17)
	t2		1 (11)		3 (23)	2 (33)
/ʁ/ → [ɾ]*	**t1**		2 (22)	3 (27)	1 (8)	
	t2		1 (11)	1 (9)	2 (15)	
/ʁ/ → [j]	**t1**				1 (8)	
	t2					
/ʁ/ → [l]	**t1**		2 (22)			
	t2					

Note: **Time**: time-point, **BPR**: biphonemic realisation, *****phonetic patterns, **1**This pattern also occurs as a dialectal variation in the Rhineland region where some of the participants of this study originated from. However, it has also been frequently reported to be a CLT phenomenon in Turkish-German bilingual speakers (Androutsopoulos, 2001; Bücker, 2007), produced by less than 15% of the age group, blank cells: pattern did not occur in this age group

In contrast to t1, children did not show any instances of phonological CLT anymore. All observable CLT phenomena were phonetic. These most frequently affected the production of the German /ʁ/, which is congruent with t1.

Developmental progress in the occurrence of CLT phenomena in German from t1 to t2

Overall, nine of the eleven children demonstrating CLT phenomena at t2 showed the same type of CLT at t1 but not always to a frequency of five times. For two children, no CLT instances were observed at t1.

<u>Turkish</u>

The types of CLT observed in Turkish at t1 either remained present in the children's speech behaviour, disappeared or increased from t1 to t2 (see Table 6.10). For example, children had overcome the production of an incorrect stress pattern, whereas the substitution of /ɾ/ by [ʁ] was observed in slightly more participants than at t1. Overall, however, CLT was relatively infrequent in children's Turkish productions and similar to t1, most frequently affected the Turkish /ɾ/. No new types of CLT were observed in the children's speech data at t2.

Table 6.10: Cross-linguistic transfer observed in Turkish by children of the t2-cohort at t1 and t2

Pattern	**Age / Time**	**4;0 - 4;5** *n* (%)	**4;6 - 4;11** *n* (%)	**5;0 - 5;5** *n* (%)	**5;6 - 5;11** *n* (%)	**6;0 - 6;5** *n* (%)
Structural simplifications						
Incorrect stress pattern	**t1**		1 (11)	2 (18)	1 (8)	
	t2					
Systemic simplifications						
/ɾ/ → [ʁ]*	**t1**		1 (11)		1 (8)	1 (17)
	t2		1 (11)	2 (18)	1 (8)	1 (17)
Vowelisation of /ɾ/	**t1**			1 (9)		
	t2			1 (9)		

Note: **Time**: time-point, *****phonetic pattern, produced by less than 15% of an age group, blank cells: pattern did not occur in this age group

Developmental progress in the occurrence of CLT phenomena in Turkish from t1 to t2

Four of the five children demonstrating CLT at t2 also showed the same type of CLT at t1 but in two cases to a lower frequency than five times. One child did not show the same type of CLT at t1 (i.e., /ɾ/ → [ʁ]) but the two other types (i.e., incorrect stress pattern and vowelisation of /ɾ/). Only one child at t2 showed evidence for a bi-directional transfer. All other children either demonstrated CLT from German to Turkish or vice versa.

6.5 Summary – Phonological acquisition at t2

The longitudinal data overall showed that children improved their phonological skills in both languages from t1 to t2 by having significantly more consonants and CCs acquired, and by demonstrating smaller numbers of PhonVar (incl. phonological patterns, InfVar and CLT) at t2. The only exception is the CC production in Turkish, which remained comparable to t1. The data further revealed that on average Turkish-German bilingual children had completed their acquisition of all consonants and CCs in German by the age of 6;0 - 6;5 years, with only a few exceptions. In contrast, children's phonological acquisition in Turkish was not completed by the age of 6;0 - 6;5 years as the children had still a few consonants and consonant clusters missing from their inventories and still had not overcome all phonological patterns from t1. PCC-A-scores significantly increased between t1 and t2 in both languages. However, children achieved on average higher scores in German than in Turkish and only one child achieved a score of 100% in German.

For all aspects assessed, a large inter- and intra-individual variability in the phonological development could be observed. Some children showed a significant improvement whereas others maintained their developmental level from t1 or scored lower than at t1. However, children's development was not identical in both of their languages. More frequently, they maintained their developmental level from t1 or exhibited a slightly weaker performance in their Turkish than in their German productions at t2. Additionally, the small sample size at t2 had an impact on the results in that developmental progression was not always observable on the group level.

7 DISCUSSION II – Children's progress in the acquisition of German and Turkish phonology over time

Based on the collected single-word naming data, the progress 43 Turkish-German bilingual children showed over a period of 12 - 15 months was analysed and illustrated. This section discusses children's phonological development over that time and specifically addresses research question 3:

> 3) How does the phonological acquisition in Turkish-German bilingual children progress over a course of 12 - 15 months?

This discussion chapter is structured into three subsections, which are in line with Discussion I (i.e., Chapter 5):

a) Differences in performances on quantitative measures from t1 to t2
b) Differences in the use of phonological patterns from t1 to t2
c) Differences in the occurrence of cross-linguistic transfer from t1 to t2 and comparison of performances in German and Turkish over time.

Longitudinal data are an important resource for the description of typical phonological development as they enable monitoring and quantifying of children's speech progression over time and allow for determining children's individual developmental profiles (Diggle, Heagerty, Liang, & Zeger, 2002). This is particularly valuable for understanding variability in cohort studies (Kirk & Demuth, 2006) and for separating intra-individual variability from clinical delays (i.e., developmental stagnations) and disorders (see especially Section 9.2.1), so that in case of a SSD intervention can be initialised early. Moreover, longitudinal data offer the opportunity to test hypotheses about how children acquire the phonological system(s) of their ambient language(s) and therefore constitute a crucial data base in the evaluation of phonological theories (Stoel-Gammon & Dunn, 1985; Vihman, 2014).

With the follow-up assessment included in the present research, it was possible to collect the first longitudinal data on Turkish-German bilingual children's phonological acquisition. These will help to gain insight into the permanence and progressions of children's phonological skills in both of their languages and will therefore make it possible to verify observations from cross-sectional data at t1 (e.g., regarding age-related changes in children's performances and inter-individual variability). Further, these data are hoped to facilitate the identification of children with an atypical phonological development, which is discussed in Section 9.2.1.

As for the discussion of t1-results, children's age of acquisition will be addressed and discussed only as the age of mastery was too sensitive to the small numbers in the age groups.

7.1 Differences in the performances on quantitative measures from t1 to t2

Based on findings from previous longitudinal research in bilingual children (e.g., Gildersleeve-Neumann et al., 2008; Gildersleeve-Neumann et al., 2009; Holm & Dodd, 2006; Morrow et al., 2014; Schnitzer & Krasinski, 1994) and the notion that maturation influences phonological development (as, for example, shown by the significant correlations between children's chronological age and their phonological skills at t1; see Section 5.1), it was hypothesised that children's performances on quantitative measures would generally improve from t1 to t2. Due to children's individual learning strategies and language acquisition circumstances, however, inter-individual variability was expected. Additionally, it was thought possible that the acquisition of new knowledge may cause a period of updating phonological systems that may present itself in form of a stagnation or regression in some children's phonological skills over time (Kim et al., 2017).

Comparisons of children's group performances at t1 and t2 revealed an overall developmental trend on the included quantitative measures (i.e., consonants and CCs acquired, PCC-A-scores and number of InfVar) over time. These could be observed in form of an acquisition of more consonants and CCs, the achievement of higher PCC-A-scores and the production of fewer InfVar in both languages compared to t1. Apart from CC acquisition data in Turkish, statistical analyses were highly significant for these improvements and supported the mean group findings. Further, group results of several measures suggested that children's acquisition was nearly complete in both languages for the oldest age group (6;0 - 6;5 years). Children at this age had acquired all German consonants and nearly all Turkish consonants, had acquired all but one onset and one coda CC in German, achieved PCC-A-scores in German that approached the maximum of 100% (range in the oldest age group: 96.5 - 100%) and attained PCC-A-scores in Turkish that ranged from 83.7 - 93.2% ($M = 88.6\%$) which is slightly higher than for t1 (range in oldest age group: 81.1 - 92.2%, $M = 86.9\%$).

These improvements over the course of 12 - 15 months on the group level confirm the expected developmental trajectory which is also reported for monolingual and other bilingual children in cross-sectional and longitudinal studies (Dodd et al., 2003;

Gildersleeve-Neumann et al., 2008; Holm & Dodd, 1999c; MacLeod et al., 2011; Morrow et al., 2014; Stoel-Gammon & Dunn, 1985; Topbaş & Yavaş, 2006). Thus, they reflect maturation as well as further language learning which is both expected with age (Davis & Bedore, 2013; McLeod, 2013; Weiss, 2007). Moreover, these long-term improvements support findings from t1, which revealed a strong correlation between children's phonological skills and their chronological age.

Despite these significant progressions, participants' group performances are still considered to be decelerated across all measures if compared to monolinguals of both languages (Ege, 2010; Fox-Boyer, 2014b; Schaefer & Fox-Boyer, 2016; Topbaş & Yavaş, 2006). As discussed for t1 results (see Section 5.1.1), this may support the notion that the additional processing and memory load bilinguals are facing slows down their rate of phonological acquisition (Fabiano-Smith & Goldstein, 2010b; Michael & Gollan, 2005; Paradis & Genesee, 1996) or may provide evidence for a negative transfer as suggested within MacWhinney's (2005a) Unified Competition Model. Following up on the latter revealed that children had not acquired all unshared consonants until the age of 5;5 - 5;11 in German and 6;0 - 6;5 years in Turkish. Their onset CCs in German were not fully acquired by 6;0 - 6;5 years either, which is significantly later than for monolinguals and thus indeed suggests a negative transfer. In contrast, there was no clear evidence for MacWhinney's (2005a) hypothesis of positive transfer in the group data, since children's acquisition of shared consonants was only completed in German but not in Turkish and their acquisition of coda CCs was still incomplete in both languages, i.e., lagging behind the rate of acquisition in monolinguals. This suggests that shared features between German and Turkish did not form more pronounced cues or were not perceived as such by all children so that they did not acquire shared features faster than monolinguals. But as for t1, a final confirmation of the model (which was not the primary aim of this thesis) will only be possible if language experience patterns are controlled for.

As for t1, however, these rates of acquisition are broadly in line with data on other bilingual children aged five to seven years (Goldstein & Bunta, 2012; Grech & Dodd, 2008; Holm et al., 1999; Lee et al., 2015b; Morrow et al., 2014; Stow & Pert, 2006; Yavaş & Goldstein, 2006). Although, children in these studies had achieved a very good level of phonological competence by about six years, they still had to acquire some language-specific consonants, fine phonetic details, and/or complex syllable structures (e.g., consonant clusters) thereafter. Findings from Lee et al. (2015b) indicate that this finalisation of phonological acquisition may still require another few years (especially if children are acquiring their languages sequentially) since their 7;0- to 7;11-year-old Mandarin-English sequential bilingual participants had still incomplete Mandarin and English phoneme inventories. Hence, it is conceivable that this will also apply to the Turkish-German bilingual participants of the current research. To prove this, however, further longitudinal data (i.e., more time-points) of this population are needed, the collection of which was beyond the scope of this research project. The comparison to other bilinguals, nevertheless indicated that Turkish-German bilingual children's rate of acquisition is typical for bilingual children with a varying age of onset of their L2.

7.1.1 Intra-individual variability in children's phonological performances over time

Besides an overall progress of phonological skills from t1 to t2, group and individual data further showed variabilities in children's development that appeared as a lack of improvement (stagnation) or a lack of maintenance (regression) of the achieved skills. Group data, for example, demonstrated an acquisition or mastery status for some consonants (e.g., /z/ in German, /z, v, g/ in Turkish) and CCs (e.g., /bʁ, tʁ, gʁ, ʃl, ʃp/ in German, /lp/ in Turkish) at t1 that could not be confirmed for t2. Individual data revealed that some children had the same or less consonants/CCs acquired, achieved lower or similarly high PCC-A-scores and/or exhibited a higher/similarly high number of InfVar at t2 than at t1. This affected either both or just one of the children's languages. Excluding children who achieved maximum scores at both time-points, these developmental trajectories were particularly interesting as they allowed for the theoretical exploration of possible influential factors on inter-individual variability; especially as cross-sectional data at t1 had also revealed inter-individual variability.

When comparing the children's longitudinal performances with those of monolinguals and other bilingual children reported in the literature, generally the same types of (large) inter- and intra-individual variabilities in children's performances could be observed (e.g., Anderson, 2004; Gildersleeve-Neumann et al., 2008; Gildersleeve-Neumann et al., 2009; Holm & Dodd, 1999c; Kim et al., 2017; Morrow et al., 2014; Schaefer & Fox-Boyer, 2016; Schnitzer & Krasinski, 1994, 1996; Watson & Scukanec, 1997). This implies that variability is a common phenomenon and not specific to bilinguals. In accordance with the reviewed literature, the scattered and intermittent consonant and CC acquisition across the included age range as well as across time-points are considered to be specifically attributable to the sensitivity of small sample sizes to inter-individual variability (see Kim et al., 2017; Munro et al., 2005; Stow & Pert, 2006). Since those instances were already evident at t1 (i.e., t1-data of the t2-cohort) it seems likely that they are ascribable to the unequal interval between the two assessment times across participants. As t2 took place 12 - 15 months after t1 this meant that participants in t2 age groups did not necessarily belong to the same t1 age group (see also Section Sample composition at t1 and t23.2.2) and may had achieved a different level of phonological competence at the date of testing. Differences in the initial acquisition status, however, may have an effect on the rate of the development as suggested for language development in school-aged children (Seltzer, Choi, & Thum, 2003) so this may have confounded the findings. Overall, the sample sizes in the age groups of the present research seem to be too small to reflect an improvement on average scores in every age group.

Comparing intra-individual performances found in the present research with those reported in studies with other bilinguals though, differences could be found. If children's individual scores at the first and last assessment time in many of the studies in the literature were compared, a clear improvement could be determined (Gildersleeve-Neumann et al., 2008; Holm & Dodd, 1999c; Morrow et al., 2014; Schnitzer & Krasinski, 1994, 1996). But this was not necessarily the case in the present research. Since assessment intervals in the cited studies were usually a lot smaller than in this research project (i.e., monthly to 8-monthly vs. 12 - 15 months), the question arises of why some children in the present research did not show any quantitatively noticeable improvement over the course of approximately one year.

Taking a closer look at children's individual performances revealed that, for example, some children had acquired new consonants/CCs at t2 but did not maintain the acquisition level of some sounds/CCs that were already acquired at t1 – at least not to the cut-offs used in this research project. Thus, it appears that the acquisition of new knowledge triggered an update/reorganisation of these children's phonological systems which resulted in a temporary loss of attained skills (cf. Munro et al., 2005; Stow & Pert, 2006). This apparent loss could evidence a temporary phase within a U-shaped developmental course (Werker et al., 2004) or demonstrate a phase of negative growth during a fluctuating developmental trajectory (Iglesias & Rojas, 2012) which have both been reported for bilingual children. Kim et al. (2017) proposed that a U-shaped learning curve, which is commonly known from much younger children than included in the present research, may occur more frequently or persist longer in bilingual children due to their necessity to specify phonemes and phonological realisation rules not only within one language but also between their languages. This implies that children's information processing systems may be (temporarily) overloaded due to acquiring two languages so that they fall back on a level of information processing they used to apply before (Werker et al., 2004). Similarly, Iglesias and Rojas (2012) argue that due to language being a complex and dynamic system that interacts with the environment, changes to any of its subsystems (e.g., phonology) may become evident as a temporary negative, positive or no growth (static development) trajectory in children's performances. Hence, an update of the system due to acquiring new knowledge (as reported further above) or a change in children's language experience patterns (which is very likely to occur over time; Hammer et al., 2011; Willard et al., 2014) may have triggered these different trajectories.

However, a stagnating or regressive developmental profile may also be the sign of an atypical phonological development, especially if it occurs in both of a child's languages (Holm & Dodd, 1999a, 1999b; Holm et al., 1999). Therefore, further analyses are required. First, it would be necessary to evaluate the speech of those children who showed a stagnation/regression in only one language separately from those whose phonological skills seemed to stagnate in both languages, since in contrast to affecting both languages, a stagnating developmental profile in only one language is generally associated with different language experience patterns in children's two languages (see Gildersleeve-Neumann et al. 2009; MacLeod et al. 2011 and Section 9.1 for a further discussion). Second, further assessment time-points are needed to compare children's performances across multiple points in time and examine their individual developmental trajectory and provide insight into the nature of the development (typical vs atypical). This would also allow for simultaneous evaluation of any changes in children's (language) environment to be able to draw any causal conclusions for influential factors (Diggle et al., 2002).

7.1.2 Possible confounding factors at t2

In addition to the possible confounding factors outlined at t1 (i.e., individual learning strategies and maturation paces, the applied cut-off criteria, methodological differences between studies, familiarity with the test items, artefacts of the test items and the quality and quantity of input children receive; see Section 5.1.2), three aspects need to be highlighted for t2-data.

First, as stressed further above, sample size issues seemed to play an even larger role at t2 compared to t1 given that the sample was only half the size of that at t1. This illustrates the importance of collecting data from significantly larger cohorts if normative data were to be aimed for (McCauley & Swisher, 1984). Second, as has been mentioned in Section 3.2.3, there were five participants who were receiving or had received SLT intervention between t1 and t2. Two of these participants were aged 4;10, the other three were aged 5;7, 5;9 and 5;11 respectively. Hence, results of the age groups 4;6 - 4;11 and 5;6 - 5;11 may be biased by this. In which way they may be influenced (i.e., positively as children improved their phonological competences with therapy, or negatively because their performances were still significantly weaker than the average typically developing child in the respective age group) will be discussed in Section 9.2.1.3 when discussing the categorisation of children as typical or atypical. Third, it is possible that the quantitative measures used were not sensitive enough to measure change in the phonological performances of all children (cf. MacLeod et al., 2011; Newbold, Stackhouse, & Wells, 2013). As will be discussed in Section 7.2, additional information from qualitative analyses appear to be necessary to complement the picture of children's phonological abilities.

7.2 Differences in the use of phonological patterns from t1 to t2

In line with quantitative performances, it was hypothesised that children would also improve their phonological skills qualitatively by specifically overcoming their phonological patterns over the course of 12 - 15 months. A follow-up assessment was further thought to offer the possibility to examine the permanence of those patterns that had just met the cut-off criteria at t1 (i.e., occurring at least 5x within an individual and in at least 15% of an age group) in the sample tested.

On the group level, similar developmental improvements as seen in quantitative data could be observed. Children had generally overcome a large number of phonological patterns from t1 to t2 (i.e., they only showed five in German and eight in Turkish) and no patterns could be observed in the two oldest age groups in German (i.e., 5;5 - 5;11 years and 6;0 - 6;5 years). Especially those patterns that had just met the cut-off criteria for being present in an age group at t1 (see Section 5.2) were mostly not produced in any language anymore. The only exception is the pattern metathesis in Turkish. It has to be noted, though, that the continuing presence of voicing and labialisation in German could not be examined as none of the children participating at t2 demonstrated these patterns at t1.

These findings first of all fully confirm the postulated hypothesis and show that further learning, such as the acquisition of more consonants and the ability of producing them in different positions in the word and varying conversational contexts, as well as maturation become especially evident in qualitative performances (Davis & Bedore, 2013; Dyson & Paden, 1983; Watson & Scukanec, 1997). The data also provide additional evidence for the overall decelerated acquisition compared to monolinguals (cf. Fox-Boyer, 2014b; Topbaş & Yavaş, 2006) which were found in quantitative data (see Section 7.1) and were also reported and discussed for t1 (see Section 5.2). Similar to t1, this decelerated acquisition was not present in all children. Some participants performed within the monolingual norm as they had overcome all phonological patterns in German by the age of five and by the age of four in Turkish (cf. Fox-Boyer, 2014b; Topbaş & Yavaş, 2006). Others even stopped using any phonological patterns by the age of four in German and therefore demonstrated an on average faster acquisition than monolingual German-speaking children (cf. Fox-Boyer, 2014b). As for quantitative data, this large inter-individual variability highlights (a) that bilingual children are in theory capable of acquiring their phonological skills within the monolingual norm (i.e., demonstrate evidence for a variation of acceleration hypothesis; Fabiano-Smith & Goldstein, 2010b, and positive transfer; MacWhinney, 2005b) and (b) that an on average slower phonological acquisition than in monolinguals does not automatically imply a clinical delay (Fabiano-Smith & Goldstein, 2010b). Especially, since children's age of overcoming phonological patterns in general is broadly in line with that reported for other bilinguals (Kim et al., 2016; Lee et al., 2015b; Morrow et al., 2014; Stow & Pert, 2006), participants' rate of acquisition on qualitative measures does not seem to be unusual for bilinguals. The same factors as outlined for quantitative data (e.g., small sample size, methodological aspects, language experience patterns, see Section 7.1) come into consideration when explaining the differences between monolinguals and bilinguals as well as the inter-individual variability.

Moreover, the data suggest that those patterns that only just met the cut-off criteria at t1 but were overcome by t2 are either only typical in a very young age of Turkish-German bilinguals (i.e., 3;0 - 3;5 years) or had occurred by chance at t1 due to the small number of participants in this age group. In either case, these patterns do not seem to be an ongoing part of children's speech behaviour. However, to aid in the clinical-decision making process – as is required in Section 9.2 – and to find out whether these patterns truly constitute typical phonological patterns in Turkish-German bilingual's phonological acquisition, longitudinal studies with larger sample sizes and further time-points are needed.

Besides an overall progress of phonological skills from t1 to t2 at the group level, results showed a large variability in children's individual performances over time. On qualitative measures, these expressed themselves in the form of production of the same number and types of phonological patterns as at t1 or the use of new types of phonological patterns at t2. This apparent stagnation/regression affected either both or just one of children's languages. Instances of this intra-individual variability on qualitative measures are generally known from other longitudinal studies in monolinguals and bilinguals (Bleile & Tomblin, 1991; Gildersleeve-Neumann et al., 2009; Kim et al., 2017; Stow & Pert, 2006; Watson & Scukanec, 1997) and challenge the premise of Dodd's (1995) Differntial Diagnosis System which involves the continuous decrease of phonological pattern use with age. They rather suggest explanations by those factors presented for quantitative data (e.g., a change in language experience patterns, reorganisation/update of children's phonological systems, atypical phonological development; see Sections 7.1.1 and 7.1.2). The use of new patterns, for example, seems to suggest a new strategy to deal with complex phonological structures along the lines of a reorganisation of children's phonological systems (cf. Holm & Dodd, 1999c) or an overgeneralisation of a phonemic contrast. Since most of these 'new' patterns, however, occurred very infrequently in children's speech and were already present as InfVar in their productions at t1, it is also plausible that this quantitative difference between t1 and t2 may have either occurred by chance (i.e., becoming evident only with this cut-off criteria) or because a covering phonological pattern was overcome (cf. James, 2001a). To identify which explanation holds true for an individual child, a more detailed analysis of individual cases would be necessary. Nevertheless, this example highlights the importance of analysing children's phonological variations both quantitatively (token) and qualitatively (pattern types) as well as in relation to productions of previous assessment times to especially prevent misinterpretations in clinical settings (i.e., evaluate the occurrence of a new pattern as a sign for atypical development although it had been used at a lower frequency before).

In general, it became apparent how important a comparison of children's performances on different measures is for longitudinal data to identify improvements that may have remained undetected if measures were considered in isolation. A decrease in the number of different phonological patterns from t1 to t2, for example, sometimes co-occurred with an increase in the number of InfVar on the individual level. These instances can still be considered as an anticipated outcome since a reduction in the number of phonological pattern tokens (i.e., an overcome of a phonological pattern) may proceed inconsistently resulting in the pattern(s) to occur less frequently than the threshold of five times but still being present in the child's speech (Stoel-Gammon & Dunn, 1985). Hence, despite the number of InfVar increased, children still showed an improvement in their phonological skills. A similar example are cases in which children showed a lack of improvement in consonant/CC acquisition and PCC-A-scores but demonstrated a significant decrease in the number of PhonVar in both languages. Children's speech had thus developed in accuracy/approached accuracy through an increase in the number of correct productions for already acquired sounds (e.g., mastery) or the number of patterns per item (e.g., *Zürafa* (giraffe) /zyɾʌˈfʌ/ → T1: [jyɾʌˈfʌ] (backing /z/ and gliding of fricatives), → T2: [ʒyɾʌˈfʌ] (backing /z/)) but children had not necessarily acquired new structures (e.g., consonants or clusters) or achieved errorless productions of the item. Given these findings, it needs to be questioned whether all of the used measures are equally sensitive to measure change in children's development. Thus, it appears that in line with Kim et al.'s (2017) argument, especially results from quantitative analyses (e.g., PCC-scores) should not be evaluated in isolation but in combination with those of qualitative analyses. This is why both of them were considered for the categorisation of children's performances in the present research (see Section 8.1).

7.3 Differences in the occurrence of cross-linguistic transfer from t1 to t2 and comparison of performances in German and Turkish over time

7.3.1 Cross-linguistic transfer at t1 and t2

Similar to quantitative measures and the types of phonological patterns, it was predicted that the number of CLT instances would reduce over time and therefore evidence the dynamic and rather transient nature of this form of interaction between German and Turkish in children’s speech (see Grosjean, 2013; Kim et al., 2017). Qualitative analyses of children’s phonetic-phonological patterns revealed that the number of CLT instances in both languages had indeed reduced noticeably from t1 to t2, in that fewer children demonstrated phonetic-phonological patterns ascribable to a CLT. This was present for both languages but was more prominent for a transfer from Turkish to German. Also, the number of different types of CLT reduced over time so that in German, for example, only phonetic CLT phenomena could be observed at t2. Nevertheless, CLT instances were present across age groups at t2 and were produced by a number of children at a lower frequency than five times.

The data of the present research therefore confirm the formulated hypothesis and support Grosjean’s (2013) as well as Kim et al.’s (2017) notion that CLT is dynamic and a process rather than a product in children’s speech. Especially its infrequent occurrence in many children and presence across age groups illustrates that it may indeed be ascribable to encoding mechanisms which can collapse under certain conditions (e.g., stress, tiredness; Grosjean 2013) rather than being permanent and related to age and maturation (Fabiano & Goldstein, 2005). Hence, this implies that CLT phenomena are to be expected in bilingual children’s productions at any point in time (if only at a low frequency). Further the data confirm assumptions from t1 regarding the limited interaction between German and Turkish and children’s ability to keep their phonological systems separate for most of the time (Fabiano-Smith & Goldstein, 2010b; Paradis & Genesee, 1996).

Beyond that, results emphasise children’s progression in phonological skills that were also found on other measures. Children seemed to have moved on from transferring sounds and rules irrespective of intelligibility losses to transferring only those sounds and rules that do not sacrifice meaning in the other language (i.e., phonetic transfer) – at least in German. This shows that although CLT might not stop occurring in bilinguals’ speech it is likely to change over time and has increasingly less systematic effects on intelligibility; although this needs to be confirmed with future research.

Performances on the individual level mainly supported findings from the group level. However, there was a small number of children that changed the type of CLT they produced at t1. One child, for example, transferred the Turkish developmental speech pattern of gliding of /ɾ/ to German (i.e., /ʁ/ → [j]) at t1. At t2, however, this child substituted the German /ʁ/ by /ɾ/. Given that this child frequently substituted /ɾ/ by /j/ in Turkish at t1 but showed full acquisition of /ɾ/ at t2, it was assumed that he was already intending to transfer /ɾ/ to German at t1 but was not able to produce it at that time and thus transferred the typical developmental pattern he was using in Turkish instead (see also Section 5.3.1). This example suggests that bilinguals may perceive and allocate (phonetically) similar sounds across their languages to the same phonemic category (cf. Fabiano-Smith & Goldstein, 2010b; Flege, 1987) and 'treat' them in a similar way. This might also explain why participants, as well as the test assistants in the present research, rarely used the Turkish allophones for /k, g, v, l/ in their Turkish productions (cf. Section 3.4). Fine phonetic differences may be ignored for the benefit of an economic way of storing and processing the phonological systems of two languages without losing meaning and intelligibility (Flege, 1987).

7.3.2 Comparison of children's performances in the two languages over time

Longitudinally, children's phonological performances in German and Turkish were assumed to improve but the strength of this improvement would be different in the two languages since language experience patterns may vary across languages and over time (Hammer et al., 2011; Willard et al., 2014).

A comparison of the phonological performances in German and Turkish indeed revealed a different progress from t1 to t2 across languages and therefore confirmed the predictions. On the group level, improvement was in favour of German, in that there were more children who had improved their phonological skills across measures in German but showed no (or only very little) improvement or a regression of their skills in Turkish than vice versa. Additionally, this number was larger than those for children who showed no noticeable improvement in both languages. In general, the rate of acquisition in Turkish lagged behind that in German and children demonstrated a larger variability on quantitative measures in Turkish. Differences in the level of acquisition of German and Turkish were already present in the children's speech at t1 with Turkish phonological skills being slightly weaker than German skills. Hence, it seems plausible that the different 'initial' acquisition status of phonological acquisition may have affected children's rate of acquisition resulting in Turkish phonological skills mainly improving but not catching up with those in German over time (cf. Seltzer et al., 2003).

As reported for t1 (see Section 5.3.2), distributed skills and a different rate of acquisition across languages are generally considered to be typical in bilingual children (e.g., Anderson, 2004; Brice et al., 2009; Bunta et al., 2009; Holm & Dodd, 1999c; Holm & Dodd, 2006; Holm et al., 1999; Kim et al., 2016; Yang & Zhu, 2010) since only very few bilinguals experience balanced proficiency across their languages (Grosjean, 1989). Those differences are likely to be linked with the proportion of language input and output patterns (language experience) as well as with proficiency in the respective languages, as has been discussed for t1-data (see Sections 5.4.2 and 5.4.3). These socio-linguistic variables, however, are not static but mostly change over time (e.g., with nursery or school enrolment), which may have affected children's phonological skills. A decrease in the intensity and frequency of language input/output in a language may have resulted in difficulties for children to maintain or enhance their (still instable) phonological skills in that language resulting in an apparent stagnation or regression of phonological skills (Montrul, 2008). For the present case, it is conceivable that German, as the environmental and educational language, increased in significance over the 12 - 15 months' period so that children developed their phonological skills in German slightly faster, which might have contributed to eclipsing the Turkish language. This assumption could be corroborated by parents' responses in the parental questionnaires at t2. Although these data were not used for statistical analyses as at t1 (because of the small sample size in each of the categories), responses still suggest a language usage pattern in favour for German for many participants (see Section 3.2.3.3). A larger number of parents, for example, indicated that their child also spoke German with their relatives and friends at t2, whereas they only spoke German to nursery nurses at t1. Item familiarity analyses conducted at t2 further support this suggestion as they revealed that children were still more familiar to the items of the PLAKSS-II than with those of the TPA (see Section 3.3.1.3). Since there were no children who improved their skills in Turkish only, the status of the German language and its importance for the children is likely to have influenced their more pronounced improvement of phonological skills in this language (cf. Pearson, 2007; Scharff Rethfeld, 2013). A more detailed evaluation of the individual relationship between the change in language experience patterns and children's phonological skills in German and Turkish would indeed be desirable but was beyond the scope of this research.

Moreover, the qualitative differences in the two languages (i.e., the use of varying phonological patterns and the asynchronous acquisition of shared consonants across languages) confirmed the assumption from t1 regarding children's use of two phonological systems (Holm & Dodd, 1999c; Paradis, 2001).

7.4 Summary and conclusions

With regard to research question 3 it can be summarised that children's speech development over the period of 12 - 15 months revealed varying developmental trajectories with the majority of children demonstrating improvements in both of their languages and across measures over time, as expected. Instances of stagnating and regressive development in either one or both of the children's languages, however, could also be observed in some children. The exact reasons for these (probably temporary) stagnations or regressions in children's development (e.g., reorganisation or update of children's phonological systems, change in language experience patterns, SLT intervention between t1 and t2, sample size effects) remain only hypothetical in this research and require further investigations, particularly more assessment time-points to monitor children's development further and to test the newly generated hypotheses. Nevertheless, this variability in developmental trajectories confirms findings from other longitudinal studies with monolingual and bilingual children and suggests that intra-individual variability is a typical phenomenon in children's speech acquisition. This contradicts especially those formalist approaches whose premise it is that speech develops linearly (e.g., Chomsky & Halle, 1968; Jakobson, 1968) and thus provides support for functionalist approaches (e.g, Davis & Bedore, 2013; Vihman, 2014).

Further, results revealed that improvements in children's phonological development did not necessarily occur in both languages and were not always detectable with a single measure. With regard to the differentiation between typical and atypical development, which is investigated in Section 9.2, this suggests the necessity to assess both of a child's languages over time and to include quantitative and qualitative phonological measures and evaluate them in combination.

Overall, several outcomes from t1 (i.e., age-related performances, children's rate of acquisition being overall decelerated compared to monolinguals but in line with bilinguals, distributed skills across languages) were also observed at t2 which implies the stability and consistency of these results in the present cohort and confirms the conclusions drawn. They further increase the likelihood of these results being generalisable to the larger population of Turkish-German bilingual children.

8 RESULTS III – Differentiation of typical from atypical phonological acquisition

Participants in the present research were recruited from nurseries and were reported to be typically developing by their parents. However, since characteristics of SSD in Turkish-German bilingual children were not known at this time, it could not be excluded that the sample also included atypically developing children. Further, children's performances in the single-word naming tests at t1 and t2 showed a large variability quantitatively and qualitatively so that it appeared necessary to further examine whether all children were indeed typically developing or whether some of them were deviating from typical development. Hence, this results chapter focuses on the categorisation of children's longitudinal phonological skills based on their single-word naming data and their performances on additional psycholinguistic assessment tasks in order to identify markers for differentiating typical from atypical development.

Results in this chapter specifically address research question 5 outlined in Section 2.3 and their presentation is structured into two sections:

1) Categorisation of phonological skills based on single-word naming data
2) Specification of the categorisation based on performances in additional psycholinguistic tasks at t2.

The first section begins with the categorisation of children's phonological performances at t1 for the whole cohort of 84 children. This is followed by the categorisation of the 43 children's phonological skills reassessed at t2 in comparison to their results at t1. Subsequently, children's performances on the additional psycholinguistic tasks (i.e., phone imitation [consonants], NWR, and consistency of non-word production) are reported in the second section. The chapter concludes with a comparison and correlation of the results with each other and the identification of potential predictors for the 43 children's performances in the single-word naming tasks.

8.1 Categorisation of phonological skills based on single-word naming data

Children's phonological patterns at both time-points were grouped as typical, delayed or deviant in each language using data from the current cohort and previous studies on monolingual children of each language for comparison (cf. Section 3.4.8). However, as results from t1 have shown, small sample sizes can impede the clear identification of patterns that are considered typical for a specific population (see Section 5.2), which in turn affects the identification of delayed and deviant patterns. Hence, basing a clinical differentiation on this marker only did not seem to be valid and reliable for the present research so that the number of InfVar was additionally used to categorise children's productions. The number of InfVar was anticipated to indicate the variability and consistency in children's speech output and would therefore, in combination with an analysis of phonological patterns, ensure that all of the child's PhonVar were qualitatively and quantitatively taken into account when formulating a hypothesis about the nature of their phonological abilities (see Section 3.4.8.2).

Based on the nature of patterns (i.e., typical, delayed or deviant) in each language as well as the number of InfVar in German and Turkish, children were subsequently assigned to three different categories (a) typically developing children, (b) atypically developing children, and (c) the group who was unable to be categorised (referred to as 'no category', cf. Section 3.4.8). Recall, the category of atypically developing children included those who demonstrated delayed and deviant phonological patterns. This research thus confines itself to a dissociation of atypical from typical development rather than enlarging upon differential diagnosis. In the two subsequent sections, results of this categorisation process are presented separately for the two time-points.

8.1.1 Categorisation at t1

Table 8.1 displays the number of children per age group who were allocated to either the typical, atypical or no category at t1. Most children ($n = 45$, 53.6%) in this research were considered to be typically developing, a few children ($n = 11$, 13.1%) were categorised as being very probably atypically developing and about one-third ($n = 28$, 33.3%) could not be allocated to either group and were thus labelled as unable to be categorised.

Table 8.1: Identification of categories based on the children's phonological performances at t1 (*N* = 84)

Age group	Typical	Atypical	No category
3;0 - 3;5	4	0	8
3;6 - 3;11	13	3	5
4;0 - 4;5	6	4	7
4;6 - 4;11	17	3	7
5;0 - 5;5	5	1	1
TOTAL	**45**	**11**	**28**
% of sample	**53.6**	**13.1**	**33.3**

Children in the atypical group demonstrated combinations of a high number of InfVar in both languages and delayed (*n* = 1) or deviant phonological patterns in one language (*n* = 1) or exhibited delayed and/or deviant phonological patterns in both languages (*n* = 8). One child was categorised as potentially atypically developing based on a high number of InfVar in both languages only (for detailed profiles of all participants see Appendix W and Appendix X). Children of all age groups were relatively equally represented in the three categories. The group of 3;0- to 3;5-year-olds was the only age group in which no children were identified with atypical development; however, this group had a proportionally high number of children who were unable to be categorised.

8.1.2 Categorisation at t2 in comparison to t1

The subgroup of 43 participants reassessed at t2 were again allocated to one of the three categories based on their phonological performance in the single-word naming tasks at t2. Figure 8.1 demonstrates the number and percentage of children per category at both time-points. Compared to the whole t1-cohort (*N* = 84; see Table 8.1), the percentages of children per category remained very similar. Children in the typical category formed the majority, whereas children whose speech was considered to be developing atypically constituted the minority. Slightly more than one-third of the children could not be categorised at t2.

The figure further illustrates that approximately half of the children (*n* = 22, 51.2%) remained in the categories they were allocated to at t1, whereas the others shifted the categories in nearly all possible ways. The greatest amount of movement occurred amongst children allocated to no category at t1; they shifted to either typically or atypically developing at t2 or remained unable to be categorised.

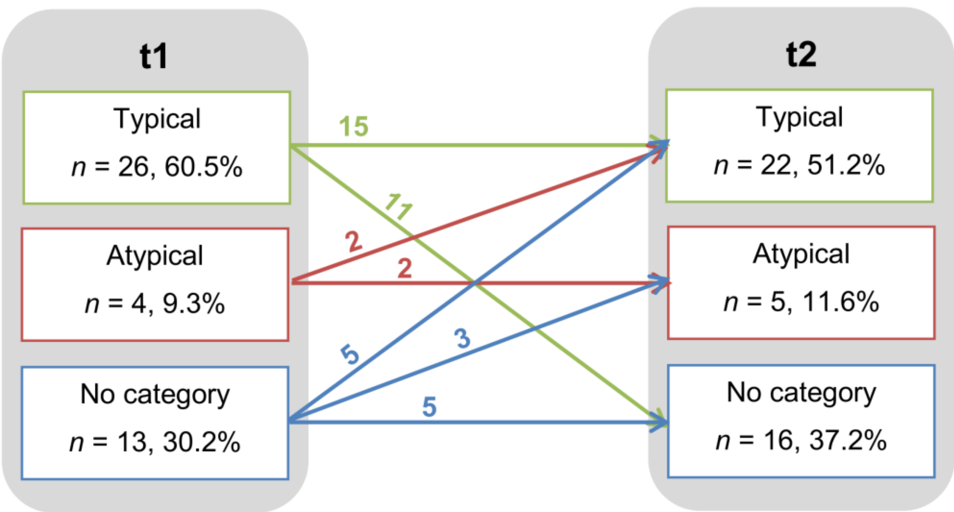


Figure 8.1: Category shift from t1 to t2 (*N* = 43)

As stated in Section 3.2.3.6, five children received SLT support between the two assessment times. Three of them remained in the categories they were allocated to at t1 (typical [*n* = 1], atypical [*n* = 1], no category [*n* = 1]). The other two children shifted the categories with one child overcoming a delay from t1 to t2 and the other child having overcome their deviant patterns in one language but demonstrating a high number of InfVar in both languages at t2 (i.e., shifting from no category to atypical).

A detailed overview of each participant's phonological profile at both times including their allocation to the different categories can be found in Appendix X. The main findings presented in this table are summarised below:

- There was an equal number of children remaining in their categories and shifting to a different category in the youngest and the oldest age groups at t2. In the older four-year-olds (4;6 - 4;11 years), 88.9% (*n* = 8) shifted the category whereas 11.1% (*n* = 1) remained. Nearly the opposite was the case for the five-year-olds. Seven children (63.6%) in the age group of 5;0 - 5;5 years remained in the same category whereas four (36.8%) shifted to another category from t1 to t2. Similarly, nine children at the age of 5;6 - 5;11 years stayed in their allocated category and four (30.8%) changed it by t2.
- Children who were considered to be typically developing at t1 but who could not be categorised at t2 predominantly demonstrated a delay in the suppression of their phonological patterns in Turkish or exhibited a large number of InfVar in Turkish. Only a few of them showed atypical performances in their German productions.
- Those children who were categorised as atypical at t1 but performed typically at t2 had overcome their phonological delays from t1 in both languages.
- Children who were deemed unable to be categorised at t1 but performed typically at t2 predominantly had experienced a phonological delay in German at t1 which they had overcome by t2.
- Those children who could not be allocated to any category at t1 and performed atypically at t2 had all overcome their deviant patterns from t1 but exhibited extraordinary high numbers of infrequent patterns in both languages and, in two out of three cases, continued to show phonologically delayed patterns.
- Participants who remained unable to be categorised at t2 had mostly stagnated in their development as they demonstrated phonological profiles resembling that of t1.

8.1.3 Summary

Considering the results of the categorisation process two main findings can be reported. First, based on their performances in the single-word naming tests, most children could be clearly categorised as either typically (majority) or atypically (minority) developing at both time-points. However, for about one-third of the children at t1 and t2 performance did not allow for a clear allocation to one of the categories. Second, about half of the participants shifted categories from t1 to t2. These children had either overcome their previous phonological difficulties or demonstrated a delay or stagnation in one of their languages at t2.

8.2 Specification of categorisation based on psycholinguistic performances at t2

In order to gain further information about the children's specific psycholinguistic abilities and to assist in differentiating typically from atypically developing children, their performances on additional psycholinguistic tasks at t2 were evaluated. These tasks included the assessment of (a) children's ability to imitate phones in isolation, (b) their ability to repeat non-words accurately, and (c) their competence to produce non-words consistently. The latter two were administered using quasi-language-independent non-word stimuli and thus provided the opportunity to collect discriminative information without drawing on the children's vocabulary knowledge.

The following section provides an overview of the reliability of the psycholinguistic tasks (potential predictor variables) and children's performances on them. The next section reports the results of the relationship analyses between the children's performances in the single-word naming and psycholinguistic tasks. Finally, a categorisation of children's performances is presented in the last subsection which takes the results from the relationship analyses into account.

8.2.1 Psycholinguistic skills in Turkish-German bilingual children

This subsection reports on the descriptive statistics for each of the psycholinguistic tasks included in the assessments at t2. The respective scoring criteria are outlined in Sections 3.4.5 (phone imitation), 3.4.6 (NWR), and 3.4.7 (consistency of non-word productions).

8.2.1.1 Phone imitation skills

In order to assess whether children's motor execution abilities in both languages were intact for an accurate speech output, they were asked to imitate Turkish and German consonants in isolation.

Table 8.2 presents the average number of consonants correctly imitated per age group. The data show that even the youngest children were capable of correctly imitating most consonants in both of their languages. In general, this skill was found to be significantly correlated to children's chronological age (in months) with more consonants being imitated phonetically correct in the older children: $r_s = .312$, $p = .041$.

Table 8.2: Average number of phones imitated correctly per age group

Age group	*N*	*M*	*SD*	Range
4;0 - 4;5	4	23.00	2.94	20.00 - 26.00
4;6 - 4;11	9	21.67	2.12	19.00 - 25.00
5;0 - 5;5	11	22.18	1.78	19.00 - 25.00
5;6 - 5;11	13	22.23	3.61	15.00 - 26.00
6;0 - 6;5	6	24.83	1.17	23.00 - 26.00

Note: Maximum score was 26 including both shared and language-specific consonants from Turkish and German

8.2.1.2 Repetition of quasi-language-independent non-words

To examine the children's ability to assemble new motor programs for unfamiliar words, they were asked to repeat 12 quasi-language-independent non-words. Their realisations were subsequently analysed regarding the number of phonemically accurate repetitions.

As can be seen from the standard deviations presented in Table 8.3 there was a large variability in children's NWR performances, which found its peak in the age group of 5;0-5;5-year-olds. Nevertheless, children's age and the number of accurately repeated non-words were weakly but significantly correlated: $r_s = .392$, $p = .009$.

Table 8.3: Average numbers of accurately repeated non-words per age group

Age group	*N*	*M*	*SD*	Range
4;0 - 4;5	4	5.50	2.06	4.00 - 9.00
4;6 - 4;11	9	4.56	2.63	1.00 - 9.00
5;0 - 5;5	11	4.73	3.33	0.00 - 10.00
5;6 - 5;11	13	5.69	1.64	3.00 - 8.00
6;0 - 6;5	6	9.67	0.94	9.00 - 11.00

Note: Maximum score was 12.

8.2.1.3 Consistency of non-word production

To examine the children's ability to produce newly assembled motor programs consecutively and thus to gain some insight into their motor planning skills, children were asked to consecutively repeat the items of the NWR-task another three times. Those repetitions, including the one during the NWR-task, were analysed regarding their degree of consistency. The child gained a score of one per two consistent repetitions (cf. Section 3.4.7).

Table 8.4 displays the average consistency scores per age group and again highlights large variabilities across age groups, especially in the younger four-year-olds (4;0 - 4;5 years) and the older five-year-olds (5;6 - 5;11 years). Overall, children's ability to produce quasi-language-independent non-words consistently (i.e., in the same way repeatedly), was moderately related to their age: $r_s = .423$, $p = .005$.

Table 8.4: Average consistency task scores per age group

Age group	*N*	*M*	*SD*	Range
4;0 - 4;5	4	25.50	7.51	17.00 - 34.00
4;6 - 4;11	9	23.78	4.18	16.00 - 30.00
5;0 - 5;5	11	24.55	4.18	20.00 - 31.00
5;6 - 5;11	13	27.08	5.19	15.00 - 34.00
6;0 - 6;5	6	31.67	1.63	29.00 - 33.00

Note: Maximum score was 36.

8.2.2 Predictions of single-word naming performances at t2 from psycholinguistic skills

Children's phonological performances at t1 and t2 were characterised by a large variability (see Chapter 4 and 6), hence it was considered important to find out how much of the variation could be explained by their speech processing skills and how much is likely to be influenced by other (not tested) variables (e.g., item familiarity, input processing). To assess the relative power of the quasi-language-independent psycholinguistic measures for possibly predicting children's performances in different single-word naming measures in both languages, correlations, hierarchical regressions, and commonality analyses were run. This subsection first presents the relationships within the potential predictor variables. Secondly, non-parametric correlations between the potential predictor variables and single-word naming skills in German and Turkish are reported, before finally results of the regression and commonality analyses are presented.

8.2.2.1 Relationships within the different psycholinguistic variables

Spearman's correlations were carried out within potential predictor variables. This allowed for an investigation of how the individual psycholinguistic skills were related to each other. Table 8.5 shows that all potential predictor variables were highly significantly correlated with each other and that relationships were moderate to strong.

Table 8.5: Spearman's correlations within the potential psycholinguistic predictor variables

Measure	NWR	Consistency	Phone imitation
NWR	---		
Consistency	.745***	---	
Phone imitation	.645***	.445**	---

Note: Strong correlation, moderate correlation, *** $p < .001$, ** $p < .01$

8.2.2.2 Relationships between the different psycholinguistic variables and single-word naming data

An exploration of the relationships between the potential predictor variables and the single-word naming outcome variables in both languages also revealed moderate to strong ($r_s = .417 - .736$) and highly significant correlations (see Table 8.6). Only phone imitation skills and the number of InfVar in German were not significantly correlated ($r_s = -.246$). Further, NWR- and phone imitation skills were stronger correlated to the children's phonological outcomes in Turkish than in German. The reverse was true for the relationship of consistency skills and phonological outcomes.

Table 8.6: Spearman's correlations between performances in psycholinguistic tasks and single-word naming assessments at t2

Measure	Language	NWR	Consistency	Phone imitation
PhonVar	German	-.680***	-.472**	-.518***
	Turkish	-.724***	-.417**	-.672***
InfVar	German	-.564***	-.591***	-.246
	Turkish	-.731***	-.567***	-.482**
PCC-A	German	.665***	.510***	.552***
	Turkish	.736***	.422**	.663***

Note: strong correlation, moderate correlation, weak correlation, *** $p < .001$, ** $p < .01$

8.2.2.3 Hierarchical regression analyses of psycholinguistic skills and phonological outcome measures

In order to establish the individual and shared contribution of different psycholinguistic variables to single-word naming competences, hierarchical regression and commonality analyses were carried out. Analysis investigating whether the data fulfilled statistical assumptions for running regressions models revealed that the potential psycholinguistic predictor variables fulfilled all of these criteria and assumptions (i.e., no perfect multicollinearity, homoscedasticity, lack of autocorrelation, normal distribution of residuals, independence and linearity of the variables; cf. Field, 2009).

Potential predictor variables were entered in a systematically varied order ensuring that each potential predictor was once entered first and once entered last. This allowed for measuring the percentage of outcome variance the respective potential predictor could account for and for calculating the unique percentage of the outcome variable it could explain, after controlling for all other potential predictors. In addition, the shared variance between the psycholinguistic variables (i.e., the variance they explain in combination with each other) could be computed. For all models those results approaching significance ($p \geq .05$ and $\leq .099$) are presented as well as these may reach significance in case a larger sample size were studied.

Figure 8.2 to Figure 8.7 present pie charts for each regression analysis illustrating the unique, shared and unexplained variance. The results of the hierarchical regression analyses are further summarised in Appendix Y.

Overall, the models were highly statistically significant for both languages:

<u>German</u>

Phonological variations: $F(3, 39) = 10.01$, $p < .001$
Infrequent variants: $F(3, 39) = 11.90$, $p < .001$
PCC-A-scores: $F(3, 39) = 9.32$, $p < .001$.

<u>Turkish</u>

Phonological variations: $F(3, 39) = 24.59$, $p < .001$
Infrequent variants: $F(3, 39) = 15.90$, $p < .001$
PCC-A-scores: $F(3, 39) = 26.33$, $p < .001$.

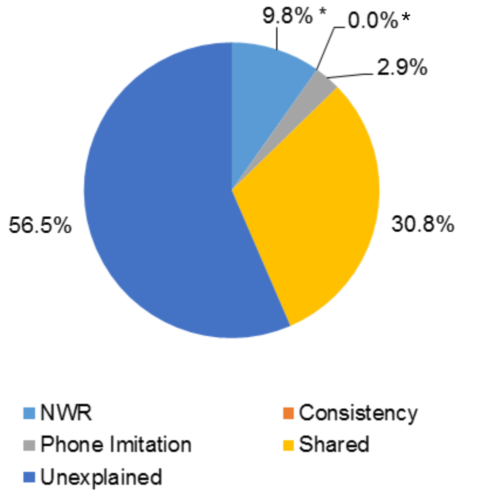


Figure 8.2: Number of PhonVar in German predicted from psycholinguistic skills

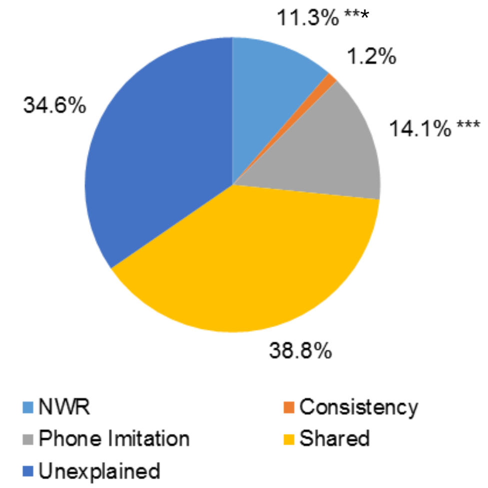


Figure 8.3: Number of PhonVar in Turkish predicted from psycholinguistic skills

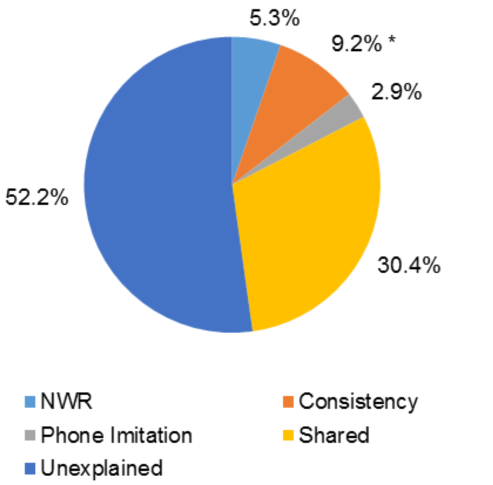


Figure 8.4: Number of InfVar in German predicted from psycholinguistic skills

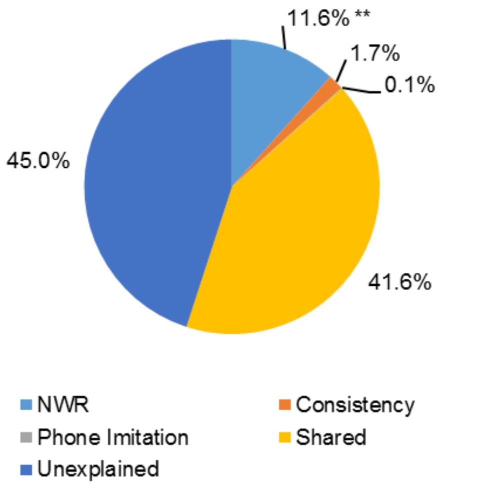


Figure 8.5: Number of InfVar in Turkish predicted from psycholinguistic skills

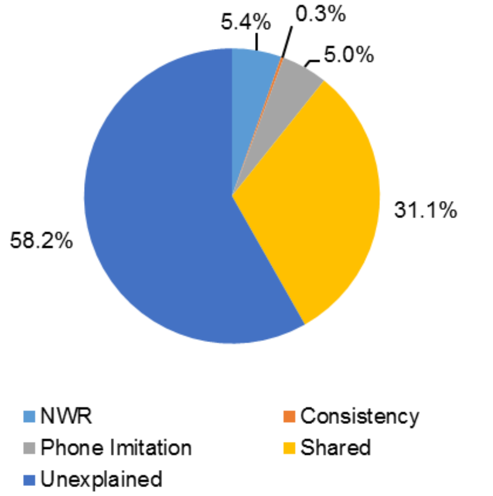


Figure 8.6: PCC-A-scores in German predicted from psycholinguistic skills

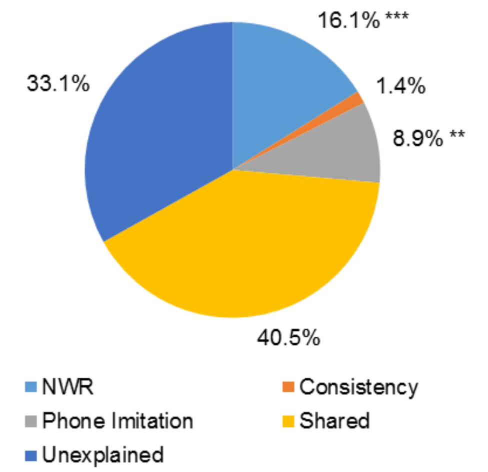


Figure 8.7: PCC-A-scores in Turkish predicted from psycholinguistic skills

Predictability of the number of PhonVar

NWR-skills uniquely explained 9.8% of the variance in the number of PhonVar in German. This contribution was significant. Phone imitation skills explained about 2.9% uniquely whereas consistency skills did not explain any part of the variance. Both these contributions were not statistically significant. The combination of all three psycholinguistic variables explained 30.8% of the variance, that is, something common to the three predictors. Overall, the model explained about 43.5% of the variance whereas the rest of the variance remained unexplained.

Results were slightly different for the number of PhonVar in Turkish. In this case, NWR-skills and phone imitation skills uniquely and statistically significantly explained respectively 11.3% and 14.1% of the variance. Consistency only explained 1.2% of the variance – which was not statistically significant. The whole model explained 65.4% of the variance which is noticeably more than for the number of PhonVar in German.

Predictability of the number of infrequent variants

For the number of InfVar in German, consistency skills were the only predictor uniquely explaining some of the variance in children's performances (9.2%) that was also statistically significant. A smaller percentage (5.3%) of the variance was uniquely explained by NWR-skills but this only approached significance ($p = .054$). Phone imitation contributed 2.9% to the explanation of the variance but was also not statistically significant. In total this model explained 47.8% of the variance with 30.4% explained by shared variance between the predictors.

Results on the number of InfVar in Turkish illustrated a completely different picture. NWR-skills were the only unique and statistically significant contributor explaining 11.6% of the variance in InfVar. Consistency and phone imitation skills uniquely but not significantly explained between 0.1 - 1.7% of the variance. Overall 55% of the variance in the number of InfVar in Turkish were explained by the model with 41.6% explained by shared variance between the predictors.

Predictability of PCC-A-scores

All potential predictor variables were found to uniquely explain some of the variance in children's PCC-A-scores in German (NWR: 5.4%, consistency: 0.3%, phone imitation: 5%), however, neither contribution was statistically significant. As a combination, they statistically significantly explained 41.8% of the variance with 31.1% of it being shared.

In contrast, NWR- as well as phone imitation skills were statistically significant and unique predictors for children's PCC-A-scores in Turkish accounting for 16.1% and 8.9% respectively. Consistency skills, however, only explained 1.4% of the variance, which was not significant. Overall, this model explained 66.9% of the variance in children's PCC-A-scores with 40.5% explained by shared variance between the predictors. This is 25.1 pp more than for German PCC-A-scores.

To summarise, children's NWR skills were found to be a unique predictor for most phonological outcome measures (i.e., the number of PhonVar in both languages as well as for the number of InfVar and PCC-A-scores in Turkish). Consistency skills were the only significant predictor for the number of InfVar in German but for no other phonological outcome measure, whereas the number of correctly imitated phones significantly predicted Turkish outcomes measures only (i.e., the number of PhonVar and PCC-A-scores). PCC-A-scores in German were not predictable by any psycholinguistic task alone. In general, psycholinguistic abilities could account for more variance in phonological skills in Turkish than in German.

8.2.3 Categorisation with additional psycholinguistic markers

Following the results of correlations, regressions and commonality analyses, children's psycholinguistic skills were taken into account when distinguishing typically developing children from those who were considered to be potentially atypically developing (cf. Section 8.1.2). In this matter, those children who could not be allocated to either group at t2 were of particular interest as it was in question whether quasi-language-independent psycholinguistic skills could support the identification of SSD.

Figure 8.8 displays the number of children per category at both time points and the respective grading of NWR- and consistency skills at t2 in the categories, (i.e., good [+]: performances ≤ 1.25 *SD* below the age group mean; weak [-]: performances > 1.25 *SD* below the age group mean; and either weak or good [+/-]: performances of either ≤ or > 1.25 *SD* below the age group mean).

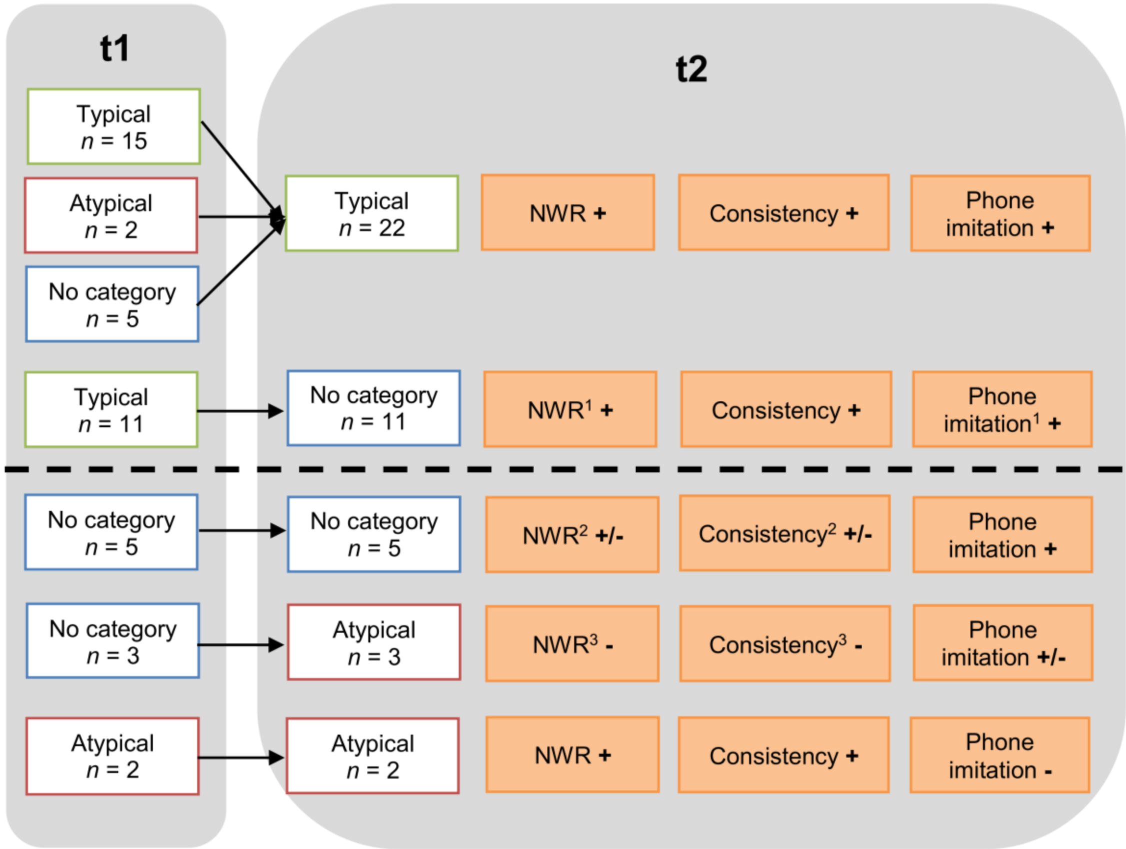


Figure 8.8: Categorisation of children at t2 including category shift from t1 and children's corresponding psycholinguistic skills at t2

Note: **+**: performances ≤ 1.25 *SD* below the age group mean, **+/-**: performances either ≤ or > 1.25 *SD* below the age group mean, **-**: performances > 1.25 *SD* below the age group mean

There were a few exceptions to the above groupings of NWR- and consistency skills:
[1]One child demonstrated NWR-skills > 1.25 *SD* below the age group mean and another child performed > 1.25 *SD* below the age group mean on phone imitation.
[2]One child demonstrated very good (i.e., above average) NWR- and consistency skills and another demonstrated normal (i.e., between age group mean and 1 *SD* below age group mean) consistency skills but borderline (i.e., between 1 and 1.25 *SD* below the age group mean) NWR-skills.
[3]One child demonstrated borderline NWR- and consistency skills.

Three main findings arose from the specification process of the identified categories:

a) Children who were categorised as typically developing at t2 all exhibited good psycholinguistic skills across tasks.
b) Children who were categorised as atypically developing at t2 demonstrated weaknesses in at least one but often more psycholinguistic predictor variables.
c) Children who could not be clearly allocated to either group based on their performances in the single-word naming tasks at t2 could be split into two groups. Group one – which predominantly consisted of children who were categorised as typical at t1 (above the dotted line in the figure) – performed well across all psycholinguistic tasks and thus resembled typically developing children. Group two – which predominantly consisted of children who did not fit into any category

at t1 (below the dotted line in the figure) – performed poorly on either the NWR- or consistency task and resembled atypically developing children. Hence these children are at an increased risk for developing SSD. For a detailed overview of children's phonological and psycholinguistic performances at t2 see Appendix X.

The figure further illustrates that each of the three psycholinguistic assessment tasks yielded important additional information on the children's phonological capabilities to not only categorise ambiguous (i.e., those that were unable to be categorised) phonological performances and identify children at risk for SSD but also to confirm findings from single-word naming assessments. Especially, NWR-skills appeared to be of discriminative power. Children who demonstrated weaknesses in this task were mainly allocated to the atypical category. Furthermore, if all typically developing children and those who were unable to be categorised but performed well in the psycholinguistic tasks were compared as a group to atypically developing children and those who were unable to be categorised with weak psycholinguistic performances (as a group), statistical analyses revealed highly significant differences in the children's PCC-A-scores of both languages, Mann-Whitney test: $U_{German} = 73.50$, $z = -2.63$, $p = .007$, $r = -.40$, $U_{Turkish} = 25.50$, $z = -4.01$, $p < .001$, $r = -.61$. Thus, combing psycholinguistic tasks with single-word naming data emerged as a promising option to distinguish typically developing children and children with/at risk for SSD.

8.2.4 Summary

Analyses revealed that the reliabilities of all three psycholinguistic tasks are mainly acceptable (Crohnbach's alpha) to very good (inter- and intra-examiner agreement) (see Section 3.3.2.3).

Regarding the children's performances on the psycholinguistic tasks, results showed that all performances were correlated to children's chronological age with older children performing better than younger children. Consonants shared between Turkish and German were imitated phonetically correctly more often than language-specific ones and children's accuracy and consistency of non-word production was depending on the number of syllables per non-word. The longer the words the less accurate and consistent were the children's productions.

Correlational analyses revealed highly significant relationships within the potential predictor variables and also between predictor and outcome variables. Overall, the psycholinguistic measures explained 41.8 - 66.9% of the variance in children's single-word naming performances. Thus, despite the significance of all models a relatively large percentage of variance (52.2 - 58.2% for German, 33.1 - 45.0% for Turkish) remained unexplained by the three predictors and must be accounted for by other factors/variables. The largest part of the explained variance was accounted for by all predictors together. Unique contributions of the single predictors were usually small. Accuracy on NWR was found to be most frequently uniquely contributing to the explanation of variance in children's phonological performances on single-word naming tasks in both languages. Phone imitation and consistency skills only did so once. In general, the models tended to explain more of the variance in children's Turkish single-word naming performances than in their German performances.

Further, children's psycholinguistic skills supported the categorisation process by highlighting those children who were at risk for SSD and at the same time confirming the categorisations of typically and atypically developing children.

9 DISCUSSION III – Differentiation of typical from atypical phonological acquisition

This chapter discusses the outcomes of the categorisation processes at t1 and t2, including the regression analyses, aiming to identify markers to best differentiate typical from atypical phonological performances in Turkish-German bilingual children. The research questions specifically addressed in this chapter are:

5) Which factors, quantitative or qualitative, can be identified to support the differentiation between typical and atypical phonological development in Turkish-German bilingual children?
   a) Can markers for a typical and atypical development be identified in actual and longitudinal phonological skills? If so, what type?
   b) Can language-independent psycholinguistic skills predict language-specific single-word naming outcomes in Turkish-German bilingual children and therefore support the identification of SSD?

Speech sound disorders in children may entail educational, social and emotional consequences, (e.g., Lewis et al., 2000; McCormack et al., 2010; McCormack et al., 2009; Nathan et al., 2004b; Preston et al., 2013) so that children at risk for/with SSD should be identified at an early age to provide them with the necessary support to mediate/reduce potential impacts. In order to make this possible, suitable assessment tools need to be available that allow for the valid, reliable and comprehensive examination of children's phonological skills (Dodd, 1995; Stackhouse & Wells, 1997; Stoel-Gammon & Dunn, 1985). Thoroughly designed single-word naming tests have been found to be an appropriate tool for this purpose and are frequently used in research and clinical practice with monolingual and bilingual children (e.g., Bernhardt & Holdgrafer, 2001; Grech & Dodd, 2008; Masterson et al., 2005; Wolk & Meisler, 1998). Within this research project, such a tool was designed for the evaluation of Turkish-German bilingual children's phonological skills in Turkish since there was previously no assessment available for this population that met international test construction requirements (see Section 3.3.1.1). A (preliminary) evaluation of the test's psychometric properties has revealed that it is reliable and also fulfils several aspects of validity (see Section 3.3.1.3). Its diagnostic validity, however, has not yet been evaluated and thus remains unclear. This is partly because clinical markers for the identification of SSD in bilinguals are not yet clearly identified and internationally agreed on. Therefore, markers for monolingual children have been mainly applied (see Section 2.2.2). Their applicability to bilingual children, however, is mostly unknown and their accessibility not always guaranteed. The required genetic and aetiological markers for Shriberg et al.'s (2010) SDCS, for example, were inaccessible from the children in this research and performances on the MSAP (the assessment battery specifically compiled for the SDCS) were not collectable due to the unavailability of a Turkish-German version of it. Therefore, the markers identified by Shriberg et al. (2010) could not be used within the present research. In contrast, information on the nature of phonological patterns, a marker used in Dodd's (1995) classification system, is generally easily accessible from a spontaneous naming sample and there is preliminary evidence for the applicability and validity of this marker with bilingual children (Holm & Dodd, 1999a, 1999b; Holm et al., 1999; Lee et al., 2015a; So & Dodd, 1994). However, as results from t1 have shown, small sample sizes may have biased the clear identification of patterns that are considered typical for this population (see Section 5.2) so that this marker was supplemented by the number of InfVar. In combination, the nature of phonological patterns and children's number of InfVar would ensure that all of the child's PhonVar were qualitatively and quantitatively considered for the categorisation process (see Section 3.4.8.2).

However, as Stackhouse and Wells' (1997) psycholinguistic framework highlights, children's speech processing can break down on many different levels of the speech processing chain. Thus, evaluating children's performances on one speech processing route (i.e., single-word naming) only may not be sufficient to identify their individual difficulties. In order to clearly differentiate between typical and atypical phonological development, psycholinguistic tasks targeting different speech processing routes were thus included in the present research and considered for the categorisation at t2 (see Section 9.2.2).

9.1 Categorisation at t1

Due to the large variability in the data found at t1 (see Chapter 4 and 5) it was felt necessary to identify whether all children of this cohort were actually typically developing. Hence, children's phonological performances on the single-word naming assessments in German and Turkish were categorised. Based on the rationale outlined above, it was predicted that a combination of the markers 'nature of phonological patterns' and 'number of InfVar' would help to identify children with typical and atypical phonological acquisition in the present research.

Results of the categorisation process at t1 revealed that the majority of children (*n* = 45) demonstrated a typical phonological development. Their phonological patterns in both languages were considered age-appropriate for this cohort and their number of InfVar was low (≤ 1.25 *SD* above age group mean) in both languages. Typical patterns and the ages for which they are considered appropriate are presented in Section 5.2. A small group of children (*n* = 11) were considered to be potentially atypically developing as they demonstrated combinations of a high number of InfVar in both languages and delayed/deviant phonological patterns in one language, exhibited delayed and/or deviant phonological patterns in both languages, or showed age-appropriate phonological patterns but a high number of InfVar in both languages. Phonological patterns identified to be deviant were: allophonic use of sibilants, contact assimilation (/tr, dr/ → [gr, gr]), nasalisation, denasalisation, epenthesis of schwa in other word positions than CCs, stopping of other sound classes than fricatives and affricates, and vocalisation of /l/ in German, as well as allophonic use of sibilants, unusual fronting, frication of stops, and tapping of the liquid /l/ in Turkish. The remaining 28 children could not be allocated to either group since their performances did not meet the criteria for either typical or atypical development (i.e., the markers ‘nature of phonological patterns’ and ‘number of InfVar’ suggested a different category for each of their languages).

These findings suggest that with the applied qualitative and quantitative markers, a clear identification of typical and atypical phonological development was possible for the majority of children. On the one hand, this provides some support for the usefulness of linguistic surface patterns for identifying SSD (Dodd, 1995; Grunwell, 1987; Stoel-Gammon & Dunn, 1985) and their application to bilingual children as suggested by Dodd and her colleagues (Holm & Dodd, 1999a, 1999b; Holm et al., 1999; Lee et al., 2015a; So & Dodd, 1994). On the other hand, however, outcomes illustrated that the nature of phonological patterns needed to be supplemented by a quantitative marker since a small number of children with potentially atypical development would have been undetected without the inclusion of the number of InfVar. These were children who showed a high number of InfVar in both languages but delayed/deviant patterns in one language and age-appropriate patterns in the other language. If the marker ‘nature of phonological patterns’ had been considered on its own, these children would have been categorised as typically developing given that they did not evidence delayed/deviant patterns in both of their languages (cf. Dodd et al., 1997; Holm et al., 1997). However, this would have resulted in disregarding children’s high number of InfVar in both languages, which in turn indicates an unusually large variability in their phonological productions. Given the large overlap in the suggested origins of variability in children with and without SSD (i.e., developmental reorganisation of the phonological system [Ingram, 1989a; Vihman 2014]; unstable and/or indistinct phonological representations [Macrae et al., 2014]; immature motor execution difficulties [Kent, 1984]; motor planning issues; [Dodd, 2005; Stackhouse & Wells, 1997]), however, it could not be excluded that the involved children were atypically developing. The facts that variability occurred in both of the children’s languages (thus fulfilling the criterion for SSD in bilinguals) and one of their languages additionally showed a delay/deviance, especially led to the assumption that these children might be at risk for SSD. It was, for example, conceivable that both of the children’s languages were affected by motor-based issues but that due to good word knowledge (Stackhouse & Wells, 1997) and/or memorised correct pronunciations for high frequent words in language A (Pascoe et al., 2006; Stackhouse et al., 2007), children’s weaknesses in language A were not detectable by an analysis of phonological patterns. Hence, a combined consideration of qualitative and quantitative aspects seems to be recommendable when aiming to identify typical and atypical development in bilingual children, a procedure that SLTs reported to generally apply to both monolingual and bilingual children albeit not with the same markers as in this research (McLeod & Baker, 2014; Skahan et al., 2007).

When considering the 33.3% of the sample that could not be clearly allocated to the typical or atypical category, however, the suggested sensitivity of the markers needs to be put into perspective. Since they suggested a different category for each of the children’s languages in this subgroup, the postulated hypothesis could not be confirmed. Showing distributed skills across languages is a typical phenomenon in bilingual children (Anderson, 2004; Bunta et al., 2009; Grosjean, 1989; Kim et al., 2016) and generally not assumed to be the sign of an atypical development (Hambly et al., 2013; McLeod et al., 2017; Scharff Rethfeld, 2013). However, studies have also revealed that speech and language difficulties may present themselves differently in a bilingual child’s languages so that a single assessment task may not be sufficient to detect atypical development in all children (Kohnert, 2010; Lindner, Mathieu, & Gagarina, 2016). Especially, a pilot study by Fox-Boyer et al. (2014) showed that the nature of phonological patterns in the speech of Turkish-German and Russian-German bilingual children with suspected SSD was not necessarily identical in each language. Therefore, it was questioned whether all children for which markers suggested a different category for each language were indeed typically developing. It was rather hypothesised that the applied markers, as well as the assessment task they were generated from, were not sensitive enough for differentiating language experience issues (i.e., the cause for most imbalances in children’s languages) from phonological deviance. This generally supports the view that children’s phonological acquisition can break down on many different speech processing levels (Stackhouse & Wells, 1997) which are not all targetable with a single assessment task. To verify these assumptions, however, further investigations of children’s performances were needed. First, it seemed to be necessary to investigate children’s phonological abilities on further speech processing routes to identify where their speech processing breaks down and whether deficits in speech processing skills can be assumed causative for the variability (see Section 9.2.2). Second, it was required to follow up these children and evaluate their further development in both languages to investigate if phonological difficulties persist in both languages so that language experience issues could be excluded. The outcomes of these investigations are discussed in the next section.

9.2 Categorisation at t2

Besides the investigation of the developmental profiles of those children whose phonological acquisition could not be clearly determined at t1, the follow-up and re-categorisation of children 12 - 15 months after the initial assessment, also allowed for a general investigation of the stability of categorisation over time. The outcomes of this second categorisation are presented in the following two sections.

9.2.1 Categorisation with phonological markers

Since group performances on the different phonological measures included in this research (i.e., consonant and CC inventories, PCC-A-scores, number of PhonVar, number of InfVar, phonological patterns) demonstrated that maturation and further language learning generally improved children’s phonological skills from t1 to t2 (see Sections 7.1 and 7.2), this outcome was expected for typical development. Based on this, it was hypothesised that children categorised as typical would remain in this category over time unless they experience a disruption in their development, triggered, for example, by a change in their language experience patterns which then resulted in a developmental delay. Children in the atypical category, however, were considered to show different developmental trajectories due to the different nature of delay and deviance (Broomfield & Dodd, 2004; Dodd et al., 1989). In line with previous research, it was suggested that children may overcome delayed phonological patterns over time (by overcoming developmental barriers), whereas children who demonstrate deviant phonological patterns are unlikely to overcome their phonological deficits without external support (Fox & Brodbeck, 2004; Zhu & Dodd, 2000a). In both cases, however, children would show some progress due to maturation and further language learning (Fox & Brodbeck, 2004).

For children who were unable to be categorised, three possible developmental paths were hypothesised: (1) some children will improve their phonological skills and demonstrate typical performances in both languages at t2 (e.g., because the imbalance in language experience patterns resolved); (2) some children will be found to be atypically developing by demonstrating atypical phonological skills in both languages over time (e.g., because good vocabulary knowledge or high familiarity with the words used in the assessments helped to mask phonological difficulties in single-word naming assessments at first but were uncovered during later development); and (3) some children will remain difficult to categorise by showing a different nature of development in their two languages (e.g., due to an unchanged imbalance in language experience patterns).

The categorisations of children's phonological performances at t2 (based on the nature of patterns and the number of InfVar) revealed different developmental progressions over time, all of which met the postulated hypotheses at least partially. That is, children who were categorised as typical at t1 continued to demonstrate age-appropriate performances in most cases ($n = 15$). The remaining eleven children showed weak skills in one of their languages at t2 and were therefore unable to be categorised. Children who were categorised as atypical due to showing delayed phonological patterns in both languages at t1 resolved their delay and performed typically at t2 (i.e., spontaneous remission of delay; $n = 2$). In contrast, children who were considered to be atypically developing due to deviant phonological patterns in both languages continued to show deviant patterns in both languages and thus remained in the atypical category ($n = 2$).

Participants who were unable to be categorised, could, as predicted, be divided into three groups of different types of developmental progressions: (1) children who performed typically at t2 (i.e., who had overcome their delayed/deviant patterns and reduced their high number of InfVar in the language that was affected at t1; $n = 5$); (2) children who performed atypically at t2 (i.e., who exhibited typical or delayed phonological patterns in one of their languages but a high number of InfVar in Turkish and German; $n = 3$); and (3) children who remained unable to be categorised at t2 (i.e., who continued to show delayed/deviant phonological patterns and/or a high number of InfVar in the language that was also affected at t1; $n = 5$). Each of the identified developmental profiles is individually addressed and discussed in the following.

9.2.1.1 Developmental profile: Remaining typical

The fact that typically developing children remained typical in the majority of cases reflects the influence of maturation and further language learning (Davis & Bedore, 2013; Weiss, 2007). Further, it supports the formulated hypothesis that children should continue showing typical performances in both languages if their development is not disrupted.

9.2.1.2 Developmental profile: From typical to unable to be categorised

In contrast to children who remained typical across time-points, children who shifted the category from typical to unable to be categorised are assumed to have experienced a change in their language experience patterns that caused variability in the productions (including delayed and deviant patterns) of their weaker/non-dominant language. This is grounded in the fact that the majority of children in this category (seven out of eleven) demonstrated a delay/deviance in their Turkish skills at t2. Given that this is not the environmental language, it is possible that its importance for children decreased over time resulting in less frequent use and practice (cf. Pearson, 2007) as well as less stable and indistinct phonological representations (Davis & Bedore, 2013). Although this assumption still needs to be verified with an analysis of children's language experience patterns at both time-points, it suggests that without monitoring these children over time the interpretation of children's membership in the 'no category' at t2 would have been less clear.

9.2.1.3 Developmental profile: From atypical to typical

This developmental profile only affected children who showed delayed phonological patterns in both languages at t1. The fact that they performed typically in both languages at t2 may imply that their development was temporarily interrupted at t1 but that they were able to overcome this disruption in the time between t1 and t2. This would be in line with findings in monolingual children, which revealed that a large percentage of children diagnosed with phonological delay have a good chance of overcoming their phonological difficulties with time and without external support if they are younger than five years (Fox & Brodbeck, 2004; Williams & Elbert, 2003; Zhu & Dodd, 2000a). A possible explanation for this is Dodd et al.'s (1989) finding that a phonological delay is not associated with a specific deficit in the speech processing chain but rather an disruption of the children's development. This disruption may be caused by a variety of internal and external factors (i.e., middle ear problems) and will be overcome as soon as these internal or external barriers have resolved (Zhu & Dodd, 2000a). Although this research did not aim for a differential diagnosis, this finding thus suggests that the involved children were indeed clinically delayed at t1. The reasons for the disruption (and the overcome of it) in their development, however, remain hypothetical within this project. It has to be noted, though, that one of these children had received SLT support between the assessment times which may have assisted their improvement.

9.2.1.4 Developmental profile: Remaining atypical

The fact that children who were initially categorised as atypical based on deviant phonological patterns in both languages actually remained in this category over time shows the stability of the applied categorisation scheme but also suggests that their phonological difficulties may be associated with a cognitive-linguistic deficit that is assumed to underlie a consistent phonological disorder in monolinguals (Broomfield & Dodd, 2004; Dodd et al., 1989). Monolingual children with a consistent phonological disorder were found to generally improve their phonological skills over time but were unlikely to overcome their cognitive-linguistic deficits without external support (Fox & Brodbeck, 2004; Zhu & Dodd, 2000a). This was also shown in the present research as children's number of different pattern types reduced over time but their frequently occurring deviant patterns remained in both languages. These similarities with monolingual children provide support for Dodd et al.'s (1997) notion that the underlying deficits of SSD are language-independent and that children's speech output is generated, restricted and managed by one single underlying mechanism.

9.2.1.5 Developmental profile: From unable to be categorised to typical

Children who performed typically at t2 but were unable to be categorised at this t1 predominantly (four out of five) exhibited atypical phonological performances (i.e., delayed/deviant patterns or a high number of InfVar) in German at t1. It was, therefore, anticipated that their initial phonological difficulties were caused by a very slow development in German possibly associated with external factors such as not enough contact time to/experience with the German language (cf. Chilla et al., 2010) or a phase of restructuring/re-organising their phonological systems (Kim et al., 2017). Similarly, for one child who initially showed atypical performances in Turkish but typical performances in both languages at t2, language experience factors (e.g., an increase in exposure to the language and in the opportunities to practise speaking it) may have played a role. This is in line with findings from Lindner, Mathieu and Gagarina (2016) who also identified a group within their 90 Russian-German bilingual children that showed atypical performances on different speech and language tasks in only one language at t1 but had resolved these by t2. The authors, therefore, concluded that language input patterns had played a role in this (Lindner et al., 2016). A comparison of individual language experience patterns at t1 and t2 in relation to their phonological performances, however, would be necessary to verify this assumption for the present cohort. This is recommended for future investigations. Nevertheless, this developmental profile demonstrated that longitudinal monitoring of children was necessary to clearly identify these children as typically developing since the markers used were not sensitive enough to allow for this interpretation already at t1. The consequences of this for clinical practice are outlined in Section 9.2.1.8.

9.2.1.6 Developmental profile: From unable to be categorised to atypical

Children with this developmental profile showed deviant phonological patterns in one language at t1 which co-occurred with a high number of InfVar in one language in two out of three cases. Although none of the three children continued to show their deviant phonological patterns, they all demonstrated high numbers of InfVar in both of their languages at t2. Thus, there was evidence that they used the phonological systems of both of their languages more variably and potentially less consistently than typically developing children at t2 (cf. Ingram, 1989b). Interestingly, evidence for these issues was already present in two of the three children's t1-data (i.e., in form of qualitative and quantitative differences to typical development in one of their languages). Hence, this outcome suggests that not only the systematic divergence from typical development but also the consistency with which this divergence occurs – irrespective of whether it is present in both or just one of the children's languages – may highlight atypical development, as Dodd (1995) and Ingram (1989b) stated for monolinguals. The atypical performance at t2 in the originally unaffected language may be a result of internal factors (Zhu & Dodd, 2000a): Children either stagnated in their speech development so that both languages appear to be affected at t2 or speech difficulties in the originally 'unaffected' language became unmasked due to having overcome other (e.g., age-appropriate) phonological patterns in that language. The latter may be the case if certain phonological patterns were blocked by others (James, 2001a). Further, it is possible that children's atypical performances were simply not (yet) identifiable with the assessment task used at t1 (see Section 9.1 for possible explanations) but that this identification was possible when their deficits became more pronounced (i.e., when children's performances differed more noticeably from that of typically developing peers at an older age) (Stackhouse & Wells, 1997).

The implications of this developmental profile are twofold: First, it challenges the widely-held view that children who demonstrate atypical performance in only one language are 'per definition' typical developing (Dodd et al., 1997; Scharff Rethfeld, 2013). Longitudinal findings of this research have shown that some of these children might actually be at risk for/have SSD but that the markers or assessment tasks to identify them may not be sensitive enough to identify atypical performances in both languages. Second, the finding shows that (until reliable clinical markers for SSD in bilinguals have been identified) children with atypical performances in only one language should be monitored over time to ensure they receive timely intervention if it is necessary.

9.2.1.7 Developmental profile: Remaining unable to be categorised

Children who remained unable to be categorised at t2 had neither overcome their atypical development present in language A nor developed difficulties in language B. Thus, the nature of their development in the respective languages stayed the same. Since the majority of children (three out of five) continued to demonstrate atypical performances in Turkish, their unchanged nature of development could be ascribed to factors such as little/decreasing contact to the Turkish language, weaker language proficiency in Turkish, or a lower importance of Turkish in their everyday lives which continues to decrease as contact to German increases (cf. Chilla et al., 2010; De Houwer, 2007; Hammer et al., 2011; Pearson, 2007; Romaine, 1995). However, it is also possible that their atypical development was not detectable in both of their languages since only one speech processing task (i.e., single-word naming) was administered (see Section 2.2).

This developmental profile evidences that neither the markers applied in this research nor a follow-up of the involved children 12 -15 months after the initial assessment added any clarity to their nature of development. Hence, it needs to be explored whether further assessment tasks (e.g., those targeting different speech processing routes) would help to finally identify a typical or atypical development in these children (see Section 9.2.2.3).

9.2.1.8 Implications of the categorisation at t2

Outcomes of the categorisation at t2 have revealed several theoretical and clinical implications. First, results suggest that a combined consideration of qualitative and quantitative markers is generally helpful for the overall evaluation and categorisation of bilingual children's phonological performances as the large number of clearly categorised children demonstrates. However, due to the large number of children who were unable to be categorised these markers do not seem to be sufficient for the identification of atypical phonological development in all children. As for t1, further testing (especially with other psycholinguistic tasks) seems to be necessary to identify whether linguistic/phonological issues exist (see Section 9.2.2.3) or whether children's apparent atypical performances have a different origin (e.g., language experience issues). Second, results have shown that a general 'wait and see' attitude does not solve the categorisation problem either (at least not for most of the children) since the percentage of children in the 'no category' almost stayed the same over time. Although, the follow-up assessment helped to identify some children as typical/atypical who initially could not be categorised (see Sections 9.2.1.5 and 9.2.1.6), this approach is little helpful in clinical practice where early identification is desirable (e.g., Fox, 2004). However, it seems to be an interim solution until reliable markers for an atypical development are identified. Assessment intervals should then ideally be reduced to three or six months to allow for an early intervention if required – particularly when children are close to school enrolment. Third, the longitudinal investigation and re-categorisation of the children has confirmed findings from t1 (see Section 9.1) which, as opposed to Dodd et al.'s (1997) assertion, show that children who only provide evidence for the marker 'nature of phonological patterns' in one of their languages are at risk for (developing) SSD. This finding was further supported by results from Lindner et al.'s (2016) study on SLI in bilingual children. These authors collected longitudinal data (t1 - t3) from 90 Russian-German speaking children on a number of speech and language skills to identify SLI. Some of their participants performed atypically in only one of their languages at t1 and a subset of them demonstrated atypical performances in both languages at a later point in time (Lindner et al., 2016). This thus suggests that Dodd's (1995) Differential Diagnosis System is to some extent insufficient for bilingual children as it does not include a recommended procedure for categorising children with atypical performances in only one language – these seem to fall through the cracks. Further, Kim et al. (2017) challenged the assumption underlying Dodd's classification system that typically developing children's phonological patterns continuously reduce over time as often suggested by cross-sectional data. Instead, the authors found a U-shaped developmental course in their 16 Korean-English bilingual children. As a consequence, Kim et al. (2017) propose a deliberate approach to using cross-sectional normative data for identifying SSD in bilingual children and rather recommend the evaluation of developmental profiles which can be supported by findings of the present research. Finally, children's category shift from t1 to t2, especially the shift from 'no category' to typical, seems to provide evidence for the strong interplay between child-internal mechanisms and environmental factors (e.g., language experience) as is suggested by functionalist approaches (see Section 1.1).

9.2.2 Categorisation with additional psycholinguistic markers

As findings of the categorisation process with the markers 'nature of phonological patterns' and 'number of InfVar' revealed, some children could not be clearly identified as typical or atypical. Therefore, it was investigated whether children's performances on three selected psycholinguistic output tasks would shed light on these children's nature of development when combined with the previously used markers. Before discussing this, however, it is first necessary to discuss the relation between psycholinguistic tasks and the single-word naming outcomes. In a second step, their value for the categorisation is explored and discussed.

9.2.2.1 Relationship of psycholinguistic tasks with outcome measures

Based on the findings in monolingual children (Vance et al., 2005) and with reference to the psycholinguistic speech processing model (Stackhouse & Wells, 1997), it was anticipated that children's performances in phone imitation, NWR and consistency in non-word production generally relate to their single-word naming performances (i.e., number of PhonVar, number of InfVar and PCC-A-scores) in both languages as all of these tasks require to some extent the same speech output processing skills.

Correlational analyses revealed that the number of PhonVar as well as PCC-A-scores in both languages were strongly and significantly correlated to NWR-skills and significantly related to phone imitation skills (strongly for Turkish, moderately for German), demonstrating a strong interplay between phonetic and phonological skills in speech production of both languages (Vihman, 2014). The finding that NWR and phone imitation were also found to be strongly and significantly correlated, further supports the likelihood that children's phonological accuracy in speech production may be associated with motor execution skills (Stackhouse & Wells, 1997). This association may work in two ways. Limited phonetic skills may constrain the phonological use of sounds as suggested by Hewlett (1985) or difficulties in input processing hinder accurate phonetic productions (Powell & Miccio, 1996). For example, if children do not perceive certain phonological contrasts they may also not produce the sound they cannot perceive accurately. It has to be noted, though, that the relationships were stronger for Turkish (strong correlations) than for German (moderate to strong correlations) suggesting that phonological outcome in German may also (in part) be influenced by other child-internal skills or external factors.

In contrast to this, the number of PhonVar and PCC-A-scores in both languages only moderately correlated with children's consistency skills, suggesting that children's performances on these measures may in part be driven by different factors. It is, for example, possible that further variables, such as language (vocabulary) knowledge, usage frequency of certain words or language experience patterns (see Sections 5.4), have influenced children's performances on the different measures to various degrees (i.e., third variable problem). However, it is also conceivable that the relationship between consistency skills and the number of PhonVar and PCC-A-scores is only present in children of a certain age. As studies with monolingual children have shown, young children's phonological productions show a larger variability than productions by older children (Stoel-Gammon, 2007; Vogel Sosa & Stoel-Gammon, 2006) but younger children's consistency skills were found to be better than their accuracy scores (Williams & Stackhouse, 2000). This relationship reversed when the children became older (Williams & Stackhouse, 2000). So the relationship between the number of PhonVar and/or PCC-A-scores and children's consistency skills in this research may also be stronger in younger children; which would require verification with future research.

Relationships between the number of InfVar (i.e., the variability in PhonVar) in both languages and consistency performances (i.e., non-word variability) were stronger than those between other outcome measures and consistency skills but were still only approaching to be strong (Turkish: r_s = -.567; German: r_s = -.591). Thus, it can only cautiously be concluded that children's variability in PhonVar (i.e., number of InfVar) and non-word variability (i.e., consistency) may be guided by the same underlying factors. Due to the type of the assessment tasks and their processing levels allocated in the psycholinguistic framework, motor planning and/or motor execution skills would come into consideration as possible factors (Stackhouse & Wells, 1997). So far, relationships between different types of variability in children's productions have only been investigated using *word variability* (i.e., the degree of inconsistency with which words were produced repeatedly within one session; e.g., Grech & Dodd, 2008; Holm et al., 2007), and *speech error variability* (i.e., the frequency of different substitutions or deletions of a target sound; e.g., Iuzzini & Forrest, 2011; Macrae, Tyler & Lewis, 2014). Word variability and speech error variability were either not significantly or only moderately related with each other (Iuzzini & Forrest, 2011; Macrae et al., 2014), leading to the assumption that they were driven by two different types of representations, semantic and phonological representation respectively (Macrae et al., 2014). The almost strong correlations between both types of variability in the present research, however, could suggest commonalities in the speech production process and indicate that both demonstrate children's stage of development towards accurate speech output (Fox, 2016). Third variable effects, however, would need to be investigated for verification.

Further, the number of InfVar in Turkish was strongly related to NWR-skills, suggesting that the two measures reflect children's phonological performances in a similar manner. Considering that especially some of the words in the TPA were not familiar to the children, it is possible that the ability to assemble new motor programs does not only have an effect on repeating non-words/unfamiliar words accurately but also on the degree of variability in speech output (cf. Section 3.3.1.3). This would be in line with Stackhouse and Well's (1997) finding that if unfamiliar words are produced for the first time, these productions are prone to errors.

As opposed to consistency and NWR-skills, the number of InfVar in Turkish only moderately correlated with phone imitation skills and in German no significant correlation could be found at all. This implies that children's degree of variability in their phonological output may be less dependent on their articulatory abilities – at least not alone – as these would probably show effects on both languages (Holm et al., 1997). Hence, as stated further above, children's variability in phonological output may be additionally associated with a variety of other factors such as misperception, representation issues, pragmatic context and trade-offs (Holm et al., 2007), language experience and proficiency factors (especially vocabulary, see Section 5.4) or a phase of re-organisation of their linguistic systems (Kim et al., 2017; Werker et al., 2004).

Overall, the large number of significant correlations between phonological outcomes in language-specific single-word naming tasks and the quasi-language-independent psycholinguistic skills confirms that accurate speech production (in bilinguals) is very complex and multi-faceted, requiring many different skills (Ingram, 1989a; Stackhouse & Wells, 1997; Stoel-Gammon & Dunn, 1985). The postulated hypothesis, however, can only carefully be confirmed since third variable problems cannot be excluded and may have especially played a role in moderate and weak correlations.

9.2.2.2 Predictive power of psycholinguistic tasks

Since SSD may be related to different underlying speech-processing deficits (Bradford & Dodd, 1996; Broomfield & Dodd, 2004; Dodd, 1995, 2011; Dodd et al., 1989; Holm & Dodd, 1999a) it was hypothesised that quasi-language-independent psycholinguistic tasks would also predict children's concurrent phonological performance on single-word naming tests in both languages. This anticipation was supported by the fact that the psycholinguistic tasks were quasi-language-independently designed (see Section 3.3.2). However, as the three phonological outcome measures are targeting different aspects of phonological output (i.e., overall accuracy/number of variations vs. degree of variability in production), it was expected that the psycholinguistic tasks would predict the individual outcome measures with different degrees of success.

Regression analyses were applied to investigate the relationships between psycholinguistic and single-word naming skills further in order to identify possible unique predictors for children's phonological accuracy in single-word naming output. These were analysed for each of the three outcome measures (i.e., number of PhonVar, number of InfVar, PCC-A-scores) separately. Results of the regression analyses are discussed by each potential predictor variable in turn.

<u>Phone imitation skills as a predictor for phonological outcome in single-word naming</u>

Results of the regression analyses have shown that phone imitation skills are a unique and significant predictor of the number of PhonVar and PCC-A-scores in Turkish at t2 but not in German. Further, the number of InfVar was not significantly predicted by phone imitation skills in either language.

The contribution of general motor execution skills, therefore, seems to be limited and the difference between the two languages may be related to the varying degree of phonetic competence children had achieved in the respective languages by the time of the assessment. Since familiarity analyses of the test items in the single-word naming assessments revealed a particularly low familiarity with items in the TPA (see Section 3.3.1.3), this might reflect that children had comparably fewer opportunities for practising Turkish speech sounds in a number of phonetic-phonological contexts necessary for imitating the phones accurately (cf. Davis & Bedore, 2013). This would resemble the effect chronological age had on the performance of stimulability tasks in monolingual children (De Castro & Wertzner, 2012; Lof, 1996). The older the children are the more practice they have had in producing the required sounds in different phonetic contexts, which is likely to have a positive influence on their abilities to imitate/produce the sound phonetically accurately (Bybee, 2001; Davis & Bedore, 2013). The fact that monolingual children as young as 2;5 years were found to already be able to imitate the phones of their ambient language accurately (Kubaschk et al., 2015) and accurate imitation skills are acquired the latest by the age of four (Williams & Stackhouse, 2000), illustrates this acquisition may be decelerated in the bilinguals' weaker language.

<u>NWR-skills as a predictor for phonological outcome in single-word naming</u>

In German, NWR-skills were the only unique and significant predictor of the number of PhonVar but they did not significantly explain any amount of the variance in the other outcome measures. In Turkish, however, NWR-skills were found to be a unique and significant predictor of all three phonological outcome measures and even constituted the only significant unique predictor for the number of InfVar.

The findings indicate that bottom-up processing skills (e.g., the ability to assemble new motor programs) are important for producing accurate phonological output in both languages if measured on a holistic measure, i.e., number of PhonVar (cf. Stackhouse & Wells, 1997; Vance et al., 2005). Hence, performances on a quasi-language-independent NWR-task could be a useful marker for identifying children with weak (and potential atypical) phonological performances.

Given that other outcome measures in German could not be predicted by NWR, however, suggests that there are other factors (e.g., language experience, language proficiency) and possibly different speech processing abilities uniquely accounting for children's degree of speech accuracy (PCC-A-scores) and phonological variability (number of InfVar) in their single-word naming output. Within the psycholinguistic framework variability and consistency in children's phonological productions are not only associated with difficulties in the ability to assemble new motor programs but also associated with motor planning skills, imprecise word form knowledge (i.e., inaccurate phonological representations) or familiarity with the items (Stackhouse & Wells, 1997) which might have had a larger impact here as well.

The fact that NWR-skills still significantly predicted children's PCC-A-scores and number of InfVar in Turkish may be associated with children's lower familiarity with Turkish than German test items. Due to the lower item familiarity (delayed) imitation was more frequently required to elicit the target words. Thus, children probably needed to assemble new motor programs and/or had to rely on their motor execution skills more often than for producing the target words in German. Further, demands on their input-processing skills and the abilities to store phonological information accurately may have been larger for the production (often delayed imitation) of unknown than known words since successful NWR relies on a variety of input and output-processing skills (cf. Snowling, Chiat, & Hulme, 1991). Hence, it would be interesting to see if the predictive power of NWR-skills changed if items of the TPA were equally well-known as those from the PLAKSS-II.

<u>Consistency skills as a predictor for phonological outcome in single-word naming</u>

Regression analyses have revealed that consistency skills were the only unique and significant predictor for the number of InfVar in German. Apart from that, however, they did not significantly predict any further outcome measure in German and none at all in Turkish.

On the one hand, these findings imply that the number of InfVar in German is indeed dependent on children's ability to produce phonological output consistently (i.e., top-down processing). Given that infrequently occurring phonological variations from adult-like speech have usually been neglected in studies and that consistency skills needed to be assessed with a separate tool, usually a 25-word-inconsistency test (Dodd, 1995; Grech & Dodd, 2008; Holm et al., 1999; Lee et al., 2015a), this finding highlights the potential usefulness of the number of InfVar for determining children's degree of consistency in German. The speech processing skills uniquely involved in the consistent production of non-words, however, are unlikely to be responsible for children's overall accuracy in their speech output in German.

On the other hand, however, this could not be confirmed for Turkish, which limits this conclusion. It is hypothesised, though, that children's familiarity with the TPA items, has again confounded the findings in that they had to rely more on their bottom-up processing abilities than their motor planning skills in order to achieve phonological accuracy and consistency. In this context, it would be interesting to analyse the predictive power of consistency skills separately for those items that children could name spontaneously and those which needed to be elicited by (delayed) imitation.

<u>Overall and shared contribution of psycholinguistic skills to prediction of single-word naming productions</u>

Together, the three psycholinguistic predictor variables (i.e., phone imitation, NWR and consistency of NWR) accounted for a significant amount of variance in each of the three single-word naming outcome variables in both languages. Shared explained variances ranged from 30.4 - 31.1% in German and 38.8 - 41.6% in Turkish. The regression models overall, (i.e., unique and shared contributions combined) explained between 41.8 - 47.8% of the variance in the single-word naming productions in German and between 55.0 - 66.9% of the variance in Turkish.

Although, this still leaves a non-negligible amount of unexplained variance per outcome measure, the large amounts of shared variances illustrate the complexity of single-word naming tasks (i.e., the incorporation of different speech processing skills needed for successful phonological performance; Stackhouse and Wells, 1997), and the overlap of skills needed to perform well on the different predictor tasks. The correlations amongst the predictor variables (cf. Section 8.2.2.1), for instance, emphasise the difficulty in disentangling the involvement of the individual psycholinguistic skills to successful phonological performance.

Further, the quasi-language-independent psycholinguistic tasks overall (i.e., unique plus shared contributions) explained a larger percentage of the variance in the Turkish productions (more than 50%) than in the German (less than 50%). As discussed previously, one explanation for this might be children's low item familiarity with Turkish test items, perhaps requiring them to rely on their motor programming and motor execution skills more frequently. If this had not been required, the percentages of variances explained might have been smaller. Given that there was still 52.2 - 58.2% of the variance in German and 33.1 - 45.0% of the variance in Turkish that remained unexplained, these need to be accounted for by additional (i.e., not assessed) factors for which a number of linguistic and psycholinguistic aspects may come into consideration. First, children's vocabulary and grammatical knowledge in each language is one reasonable, additional candidate for predicting children's phonological performance. The interdependence of linguistic domains in children's speech acquisition process is well-known from studies with monolingual toddlers (Roulstone et al., 2002; Smith et al., 2006;

Stoel-Gammon, 2011; Vihman, 2014) and has also been confirmed for bilingual children (Core & Scarpelli, 2015). Further, the analysis of influential factors on children's phonological performances at t1 revealed that children's language proficiency (as estimated by their parents) seemed to have an effect on their speech production (cf. Section 5.4.3). Second, it is likely that other psycholinguistic skills better, or to a larger degree, explain children's phonological performances. Within the present research only three different speech output processing skills were assessed but Stackhouse and Wells (1997) argue that a child's inability to realise phonologically correct words may not only result from weak output skills but also from inadequate speech input processing or imprecise storing of phonetic-phonological information. Therefore, auditory discrimination and mispronunciation detection tasks that tap into speech input processing skills and children's underlying representations (Stackhouse et al., 2007) could be useful predictor variables to investigate in future projects.

In general, results of the regression analyses only partially confirmed the postulated hypothesis as phonological output was variously predicted by the psycholinguistic skills across languages and also because not all psycholinguistic skills significantly contributed to explain the variance in the different outcome measures. Nevertheless, children's variability in phonological output measures could be significantly predicted by the combination of assessed psycholinguistic skills, suggesting that they may serve as additional markers for the identification of atypical phonological development (see next section). Given the large shared variances, however, it cannot be fully confirmed that the psycholinguistic tasks predicted the individual phonological measures differently well.

9.2.2.3 Value of psycholinguistic tasks for categorisation

Given the outcomes from some SLI studies showing that quasi-language-independent NWR-tasks seem to have sufficient diagnostic power in bilingual children (cf. Boerma et al., 2015; Chiat, 2015), children's performances on this predictor variable were hypothesised to support the differentiation between typical and atypical phonological development in Turkish-German bilingual children. It was, however, unclear whether this would also apply to the other psycholinguistic tasks and to SSD. Thus, this section now discusses the potential improvement of categorisation by additionally taking into account children's performances on the predictor variables.

When adding children's performances on the quasi-language-independent psycholinguistic tasks categorisations of children's phonological performances showed that, as expected, all children with typical development also exhibited phone imitation skills, NWR-performances as well as consistency scores within the 'normal' range (i.e., ≤ - 1 *SD*). In contrast, all of the children with atypical phonological performances demonstrated borderline performances (i.e., > -1 *SD* but < -1.25 *SD*) or weaknesses (i.e, ≥ -1.25 *SD*) in at least one psycholinguistic task. Children who were unable to be categorised at t2, however, could be separated into two groups: Those who demonstrated difficulties (borderline or weak performance) with at least one of the psycholinguistic tasks (i.e., resembling children in the atypical category) and those who performed within the normal range on all psycholinguistic tasks (i.e., resembling children in the typical category). These results imply that quasi-language-independent psycholinguistic tasks could be used as an additional marker for atypical phonological performances in Turkish-German bilingual children, as is in line with monolingual children (e.g., Dodd et al., 1989; Stackhouse & Wells, 1997; Williams & Chiat, 1993). The postulated hypothesis could therefore be confirmed.

Taking a closer look at those children whose performances were unable to be categorised with the phonological markers (i.e., nature of phonological pattern and number of InfVar) at t2, the following trend could be observed: If children had shifted to this category from typical development at t1, their psycholinguistic performances were predominantly within the normal range and suggested an overall typical phonological development. Their phonological difficulties at t2 (i.e., delayed/deviant phonological patterns or high number of InfVar in one language) were thus ascribed to a change in their language experience patterns as is in line with Lindner et al. (2016).

In contrast, if children had remained in this category since t1 and, therefore, had continued showing delayed/deviant phonological patterns or a high number of InfVar in one of their languages, their psycholinguistic tasks were borderline or weak in at least one of the psycholinguistic tasks. This performance resembled that of children categorised as atypical and points to a cognitive-linguistic cause rather than a language experience issue causing the disruption/stagnation in their development (i.e., the atypical performance in one language). This suggests that children with (continuing) atypical phonological skills (i.e., delayed/deviant phonological patterns and/or a high number of InfVar) in one of their languages may be at risk for SSD. As a consequence, the statement that a SSD will always present itself in both of a child's languages (cf. Dodd et al., 1997; Holm et al., 1999) would have to be phrased more precisely by saying that the SSD needs to be present in both languages but that the actual difficulties may be become present on different speech processing levels in the two languages.

When regarding the discriminative power of each psycholinguistic variable individually, it becomes apparent that only in a few cases atypical phonological performance was related to difficulties in phone imitation. Three of the five children with atypical performance[13] as well as one out of the 16 children who were unable to be categorised at t2 exhibited difficulties in phone imitation (see Appendix X). Difficulties in imitating speech sounds accurately could be related to a number of factors such as perception difficulties, visibility of speech-sound production or oro-motor difficulties to name just a few (Lof, 1996; Powell & Miccio, 1996). In order to identify the underlying cause, it would be necessary to check whether the phones that could not be imitated accurately in isolation are also affected in phonemic contexts (i.e., examine stimulability in its original sense). This could shed light on whether the child has correct underlying phonological representations (Tyler & Macrae, 2010) and is likely to acquire these phones by maturation or whether they would require specific intervention (Miccio et al., 1999; Powell & Miccio, 1996). Due to the small number of children with weak phone imitation skills it can be assumed that most children with atypical phonological skills (either in one or in both languages) in the present sample exhibit evidence for accurate motoric movement abilities necessary to articulate words accurately (Lof, 1996). Given these outcomes as well as the findings from monolinguals that phone imitation is an early acquired skill (Kubaschk et al., 2015; Williams & Stackhouse, 2000), it can be suggested that this psycholinguistic task may have only a limited potential to discriminate typical from atypical phonological development in bilingual children of the age of 4;0 - 6;5 years.

In contrast to phone imitation, NWR was the psycholinguistic task in which children most frequently showed borderline/weak performances (i.e., three out of five children who were categorised as atypical, three out of five children whose performances could not be clearly categorised at both assessment times, and one out of eleven children who were typical at t1 but unable to be categorised at t2). It can, therefore, be suggested that the majority of children with atypical performances in at least one language had difficulties with bottom-up processing (e.g., motor programming). Hence, NWR-skills seem to be a good marker for the present age group to identify children at risk for/with SSD. The potential role of quasi-language-independent NWR-skills to serve as a marker for SSD in bilingual children is supported by their discriminative power in differentiating between bilingual children with and without SLI (Boerma et al., 2015; Chiat, 2015; dos Santos & Ferré, 2016). The fact that NWR-skills seem to have a similar discriminative power for speech and language difficulties in bilinguals supports the notion that there are some more general cognitive-linguistic skills, such as phonological working memory, involved in successful NWR, which are not only important for accurate speech but also language production (Chiat, 2015; Gathercole, 1995).

Consistency skills were found to be borderline/weak in more than half of the children with atypical performance at t2 as well as less than half of the children with unable to be

[13] For two children this was the only weak performance in the psycholinguistic tasks, for one child it was one out of three.

categorised performances at both time points. Difficulties in the consistency task mainly co-occurred with weaker NWR-skills. This leads to the assumption that children either had difficulties with the shared processing demands of these tasks or demonstrated weaknesses in the processing skills that are unique to each of the tasks. In general, it demonstrates noticeably weak speech processing skills in more than one psycholinguistic task, which probably are a sign of SSD (Dodd, 1995; Dodd et al., 1989; Stackhouse & Wells, 1997).

In line with the psycholinguistic framework (Stackhouse & Wells, 1997), children in the present research were considered to have different strengths and weaknesses in their speech output processing chain that can influence their speech production in single-word naming tasks. Despite the focus not being on identifying the specific underlying deficits of children's phonological abilities and difficulties, as is the goal in the speech processing model (Stackhouse & Wells, 1993, 1997), the psycholinguistic assessment tasks used in this research suggested cognitive-linguistic deficits in a number of children which confirmed categorisations with the phonological markers. Moreover, the quasi-language-independent psycholinguistic tasks helped to unmask speech processing difficulties which would have not been detectable by the phonological markers identified in single-word naming tasks alone and hence identified children at risk for SSD. The results further suggest the presence of several subtypes for classification (Broomfield & Dodd, 2004; Dodd, 1995; Dodd et al., 1989; Holm & Dodd, 1999a; Holm et al., 1997; Holm et al., 1999). Within the current research project, however, those can only be hypothesised. To confirm these assumptions with a larger sample, the assessment of receptive and expressive phonological skills as well as a comparison of children's performances on different speech processing tasks, would need to be included in the assessments (Stackhouse et al., 2007).

Thus, by combining markers and principles from two 'classification' approaches (i.e., the Differential Diagnosis System and psycholinguistic model) and adding a clinical marker which has not been considered in previous studies with bilinguals before, the application of the present categorisation scheme with bilingual children seems promising. The fact that a variety of clinical markers appears to be necessary for an 'accurate' identification of children with SSD confirms findings regarding markers for SLI in bilinguals (Kohnert, 2013). Nevertheless, there are a few aspects that should receive closer attention in future investigations. First, it is unknown whether children's psycholinguistic skills would have also predicted their phonological performances at a younger age. If the level of difficulty of the applied tasks were also suitable for bilingual children aged two or three years, this would facilitate the desired early identification of children with potential SSD. Second, a comparison of the diagnostic accuracy of quasi-language-independent psycholinguistic tasks with that of language-specific ones is desirable to determine best practices in bilingual assessments. This is especially interesting since studies on the identification of SLI in bilinguals have revealed heterogeneous results (Boerma et al., 2015; dos Santos & Ferré, 2016). Fourth, studying a larger sample size would offer the possibility to review the identified nature of phonological patterns (i.e., age-appropriate, delayed, deviant) for the studied age groups as well as allow for a review of the stability and generalisability of the present categorisation scheme to the wider population. This seems especially important since the percentage of children categorised as potentially atypically developing after the application of the psycholinguistic markers is currently comparably high (i.e., 23.3%) with respect to international studies on monolinguals (Broomfield & Dodd, 2004; Campbell et al., 2003).

9.3 Summary and conclusion

In this chapter, the different steps of categorising the phonological performances of Turkish-German bilingual children were discussed.

With regard to research question 5a), the findings of the present research highlight the presence of more than one clinical marker for typical and atypical phonological development in bilingual children (i.e., the combination of children's nature of phonological patterns and their number of InfVar). These qualitative and quantitative markers allowed for the identification of (a)typical development in most children, although not in all. This may be due to single-word naming tasks not targeting all speech processing levels that could possibly break down in children's phonological acquisition (Stackhouse et al., 2007; Stackhouse & Wells, 1997). Further, it could be found that monitoring children's phonological skills over time does not only provide insight into their individual developmental trajectories but may also support the categorisation process. Especially for children who were unable to be categorised the re-categorisation 12 - 15 months after t1 predominantly shed light on the nature of their phonological development.

Regarding research question 5b), results suggest that, if combined, the included quasi-language-independent psycholinguistic tasks were able to predict children's phonological skills in single-word naming tasks. Additionally, there was some evidence that psycholinguistic tasks were unique predictors for the different speech output tasks (number of PhonVar, PCC-A-scores, number of InfVar). However, this could only rarely be evidenced in both languages. Psycholinguistic tasks further helped to identify typical and atypical development in children who were unable to be categorised with phonological markers only. In this respect, this research has shown that especially children who only demonstrate atypical phonological performances in one language need to be monitored as they may be at risk for developing/having a SSD.

Overall, the categorisation scheme designed and investigated within the scope of this research presents a promising applicability to bilingual children's phonological development and integrates well into ongoing debates about best practices for identifying SSD and SLI in bilingual children (Armon-Lotem, de Jong, & Meir, 2015; McLeod et al., 2013; McLeod et al., 2017).

10 GENERAL DISCUSSION AND CONCLUSION

This research sought to identify markers for typical and atypical phonological acquisition in 3;0- to 5;5-year-old Turkish-German bilingual children living in Germany. This final chapter brings together the findings discussed in Chapters 5, 7 and 9 highlighting the main points and addressing the over-arching research aims outlined in Section 2.3. Strengths and limitations of this research as well as its theoretical and clinical implications will be discussed before concluding with directions for future research.

10.1 Review of the over-arching research aims and findings

Despite an increasing body of research on bilingual phonological development over the last decades (see Hambly et al., 2013 for a review), information on how development proceeds in typically developing children, which characteristics it shows and how to best identify atypical phonological development in a population as heterogeneous as bilingual children, is still very limited (McLeod et al., 2013; McLeod et al., 2017); especially, if considering a specific language combination such as Turkish-German. Therefore, the present research aimed to:

1) Describe the typical phonological acquisition in Turkish-German bilingual children
2) Differentiate typical from atypical phonological development in Turkish-German bilingual children.

To what extent these aims could be achieved with the data collected and analysed in this research is summarised in the following two subsections.

10.1.1 Describing the typical phonological acquisition in Turkish-German bilingual children (Aim 1)

The present research has shown that the phonological skills in 3;0- to 5;5-year-old Turkish-German bilingual children are still developing in both languages. About a year later, significant improvements in children's phonological skills could be observed that pointed to a trend for completing the acquisition by about 6;0 - 6;5 years in both languages, with German skills being slightly ahead of Turkish skills throughout age groups. This overall slow rate of acquisition compared to monolingual children is generally associated with the extra processing and storage load bilinguals have to face due to acquiring two phonological systems at the same time (Michael & Gollan, 2005; Paradis & Genesee, 1996). However, since inter- and intra-individual variability were large for both acquisition rates in German and Turkish, environmental factors were assumed to play a significant role in children's rate of acquisition as well. This could be supported by the outcomes of an analysis of influential factors (see Section 5.4) and is in line with functionalist approaches who attribute the interconnectivity of child-internal mechanisms and external factors a great importance in children's phonological acquisition (e.g., Davis & Bedore, 2013; Menn et al., 2013; Vihman, 2014). Besides the inter- and intra-individual variability in the rate of acquisition, environmental factors (e.g., children's language input and output patterns, language proficiency) also came into consideration when accounting for qualitative differences (e.g., type and frequency of CLT, types of phonological patterns) in the phonological skills between individual children. With these findings, the current research confirms outcome from previous studies (e.g., Goldstein et al., 2010; Law & So, 2006; MacLeod et al., 2011) and highlights the importance of analysing and evaluating bilingual children's phonological acquisition always for both languages (Yavaş & Goldstein, 1998) and with regard to children's individual environmental factors (e.g., Scharff Rethfeld, 2005, 2013). This is crucial for both theoretical research and clinical practice. What remains unknown, however, is the weighting of these influential factors in children's development (both for concurrent as well as longitudinal performances).

Further, this research has revealed that bilingual children's phonological acquisition is proceeding non-linearly showing progression, regression and stagnation over time. This non-linearity is known from monolingual and other bilingual children's development, with the exception that it is usually observed in considerably younger children (i.e., one- and two-year-olds; e.g., Keshavarz & Ingram, 2002; Schaefer & Fox-Boyer, 2016; Watson, 1997). In line with Kim et al. (2017), it is suggested that bilingual children show these fluctuating developmental trajectories more frequently or for a longer period than monolinguals, as they need to specify phonemes and phonological realisation rules not only within one language but also between their languages. Additionally, it seems possible that external factors (e.g., language experience patterns) contribute to these fluctuations as well. Nevertheless, these different developmental profiles should be expected when examining bilingual children's phonological development longitudinally – also in children aged 3;0 - 6;5 years.

Overall, aim 1 of this research was considered to be achieved. It needs to be acknowledged, though, that this description of the typical phonological acquisition in Turkish-German bilingual children could only be made with regard to the measures applied and the age range tested.

10.1.2 Differentiating typical from atypical phonological acquisition in Turkish-German bilinguals (Aim 2)

Within the present research, three potential quantitative and qualitative markers to differentiate typical from atypical phonological development in Turkish-German bilingual children could be determined: (a) nature of phonological patterns, (b) number of InfVar, and (c) psycholinguistic skills (i.e., NWR- and consistency performance of quasi-language-independently designed stimuli, phone imitation). Outcomes of the different steps of the categorisation process have highlighted, though, that only a combined consideration of all identified markers supported the clinical decision-making process. This included monitoring if children changed the categories over time or not. These findings demonstrate the complexity of (Turkish-German) bilingual children's (a)typical phonological development and challenge Dodd et al.'s (1997) argument that the nature of phonological patterns may be sufficient to clearly differentiate between typical and atypical phonological development in bilingual children. Further, the finding is in line with the longitudinal data in this research (see Section 7.4), which pointed to the necessity to consider several phonological measures to clearly identify improvements in Turkish-German bilingual speech performances over time.

The necessity to combine the identified markers to yield a clear differentiation of typically from atypically developing children also emphasised that a single assessment task – as, for example, the frequently used single-word naming task – is not sufficient. This corroborated findings from research with bilingual children with SLI (Kohnert, 2013), and generally supported Stackhouse and Wells' (1997) assertion that due to the different underlying deficits of SSD children's speech processing skills need to be assessed with different tasks targeting different speech processing routes to clearly identify children's strengths and weaknesses.

Hence, overall the investigation of the differentiation of typical from atypical phonological development demonstrated that (a) different types of speech processing tasks need to be incorporated in the assessment, (b) children's phonological performances should be measured at several time-points to monitor the course of development and, for example, identify unusual stagnations, but also that (c) normative data (from a large sample) for the bilingual population under study seem to be unavoidable for a clear identification of atypical phonological performances since all markers required the availability of a norm.

Given these findings, aim 2 of the present research can be considered achieved. The applied categorisation scheme appears promising regarding its applicability with bilingual children but will require verification and refinement with larger cohorts.

10.2 Evaluation of the present research

In contrast to previous research on Turkish-German bilinguals (Salgert et al., 2012; Ünsal & Fox, 2002), the design of this research project offered the opportunity to examine a comparable large number of bilingual children (*N* = 84), which allowed for some more confident generalisations than with the pilot studies. Moreover, it allowed for the longitudinal investigation of the phonological performances of a large subgroup of 43 children, which provided first insight into the range of different developmental profiles among Turkish-German bilingual children. In addition, children's phonological performances on psycholinguistic tasks were examined regarding their power to predict children's phonological skills in single-word naming data, which had only been investigated for monolingual children to date (e.g., Vance et al., 2005). Furthermore, this is the first research project that investigated potential clinical markers for SSD in Turkish-German bilingual children that did not exclusively base its results on comparisons with monolingual norms (cf. Fox-Boyer et al., 2014; Tugay & Schultz-Ünsal, 2013). These included markers which were known to be sensitive for atypical development in monolinguals (i.e., nature of phonological patterns, NWR-skills, consistency skills, phone imitation skills) but also those that have not been investigated for monolinguals or bilinguals before (i.e., number of InfVar). Since the stability of these markers could be explored longitudinally, this increases the generalisability of these findings. Therefore, the present research extends our knowledge of potential clinical markers for SSD in Turkish-German bilingual children and points out new possible pathways in identifying atypical phonological acquisition in bilingual children in general.

10.2.1 Strengths and limitations of this research

The present research provided the possibility to revise and extend a previously designed phonology assessment tool, the TPA (Salgert, 2011), for examining Turkish phonological skills in Turkish-German bilingual children. A design of a new tool was necessary since the selection of phonological assessment tools for the Turkish of Turkish-German bilingual children living in Germany was sparse and lacked some important international test construction criteria for phonology assessments in general and bilingual children in particular (e.g., linguistic criteria, psychometric criteria; see Section 2.1.3). The TPA was geared to international test construction criteria and was preliminary examined regarding its validity and reliability within this research. Hence, the first aspect this research adds is a practical one – the availability of a new and thoroughly designed assessment tool for Turkish phonology skills.

By assessing 84 3;0- to 5;5-year-old Turkish-German bilingual children’s phonological skills in both languages, a large data base for the typical phonological acquisition in this population could be created, which allowed for a detailed description of the acquisition of consonants, CCs, phonological accuracy scores (PCC-A-scores) as well as a description of PhonVar from adult-like speech (i.e., phonological patterns and InfVar). The outcomes offered valuable insights into the specific characteristics of the typical phonological acquisition in this language pair and provide a good basis for comparison when identifying children with/at risk for SSD, which was not available from the previous pilot studies (i.e., Salgert et al., 2012; Ünsal & Fox, 2002).

The design of the present research further made it possible to extend the current knowledge base on Turkish-German bilingual children’s acquisition of phonology by collecting the first longitudinal data for this population in 43 children. Longitudinal data offer a valuable insight into the degree of progress a child exhibits over time and thus into their shape of development (Kim et al., 2017; Morrow et al., 2014). The identified large inter- and intra-individual variability in the longitudinal data confirmed findings from cross-linguistic studies (e.g., Gildersleeve-Neumann et al., 2008; Holm & Dodd, 1999c; Kim et al., 2017; Morrow et al., 2014; Schnitzer & Krasinski, 1994, 1996) but at the same time illustrated how much cross-sectional data are dependent on the child’s stage in development during the time of testing (Kim et al., 2017). Hence, with the availability of longitudinal data it is now possible to put cross-sectional data into perspective.

Other strengths of this research are the use of non-words to assess further speech processing skills as well as their quasi-language-independent design. Non-words had the advantage that children did not need to access their phonological (and lexical) representations to process them (Stackhouse & Wells, 1997) so that the influence of potential representation issues on children’s speech performances could be kept to a minimum. Beneficial of the quasi-language-independent design was the time-efficient administration as only one task (instead of two language-specific ones) needed to be administered, and, more importantly, that the stimuli could be specifically tailored to the language combination of Turkish and German. This was important since the performance in NWR-tasks can be influenced by language knowledge (Boerma et al., 2015; Summers et al., 2010).

Finally, analysing and evaluating not only the phonological patterns children produced but also the number of InfVar that occurred in children’s speech provided quantitative and qualitative insights into children’s phonological skills and revealed a new clinical marker for SSD in Turkish-German bilingual children (see Sections 9.1 and 9.2).

Besides these strengths some limitations of the present research need to be acknowledged. First, the iterative process in which the design of this research was developed made it not possible to assess participants with the additional psycholinguistic tasks from the start. This, however, would have been a possibility to verify the categorisation chosen over time and would have offered the opportunity to explore the predictability of psycholinguistic skills on single-word naming performances in a larger sample (i.e., the 84 participants from t1). Additionally, the iterative process limited the number of participants for t2 as parents and nurseries did not know from the start that an availability for t2 would be ideal. Hence, the design had an impact on the number of children that could be re-recruited for t2.

Second, as pointed out throughout this thesis, 84 bilingual participants is a large sample compared to most internationally conducted studies with bilingual children (e.g., Fabiano-Smith & Barlow, 2010; Gildersleeve-Neumann et al., 2008; Goldstein & Washington, 2001; Hui, Wells, & Howard, 2015; Lin & Johnson, 2010; Morrow et al., 2014). However, this size is too small to function as a reliable normative basis as became especially evident for the nature of phonological patterns (see Section 5.2). The uneven and partly very small numbers of children per age group at both time-points, therefore, limit the study's statistical power and constrain its generalisations (McCauley & Swisher, 1984).

Third, due to the time limits and scope of this research not all of the potentially influential factors on children's phonological development described in the literature could be assessed and analysed. Especially, no information was obtained about the quality of language input. Information about parents' language proficiency in both languages or even audio data on their German and Turkish pronunciations could have facilitated the evaluation of children's speech output in both languages (cf. Khattab, 2006). A limitation related to this, is the question format used to gain information about children's amount of language input and output in German and Turkish. As outlined in Section 3.2.3.3, parents seemed to have difficulties in answering these questions accurately, so that future research may use a different question format which facilitates the collection of these relevant information.

Fourth, every study's results are limited by the assessments and measures applied. In the present case, only assessment tasks with single-word/single-phone productions were applied so that generalisations to connected/spontaneous speech may be limited and relationships with speech input processing skills could not be investigated. Further, despite a very careful selection process for the words to be included in the TPA, familiarity analyses demonstrated that a large number of items could not be spontaneously named by all children. This also applied to the items of the PLAKSS-II, although to a lesser extent. This demonstrates the large variability and difficulty in predicting bilingual children's vocabulary knowledge across their languages and may explain the large variability in the results (see especially Section 5.1.2). Hence, future investigations should consider the adjustment of the vocabulary included in these assessments – e.g., design a separate version of the PLAKSS-II for Turkish-German bilingual children. Another aspect related to limitations in the measures is the fact that only one production per CC was used for analyses. This way generalisability was impeded (cf. Schaefer & Fox-Boyer, 2016).

10.2.2 Theoretical and clinical implications of the findings

The outcomes of this research yielded some theoretical implications for phonological development in monolingual and bilingual children as well as several implications for the clinical practice – especially the identification process of SSD – with (Turkish-German) bilingual children. These are subsequently addressed.

First, as discussions of the data have shown (see Chapter 5 and 7), formalist approaches did not seem to be able to explain all of the characteristics found in Turkish-German bilingual children's speech development. These included especially the asynchronous acquisition of shared consonants across German and Turkish, the influence of language experience patterns (input and output) on children's phonological performances, and the variability in children's performances over time (i.e., progression, regression, stagnation). However, since child-internal mechanisms are the same for children acquiring one or more languages and both monolingual children and bilingual children are influenced by environmental factors (e.g., language input; Davis & Bedore, 2013), a reliable and valid phonological theory should be able to explain both monolingual and bilingual phonological acquisition. Functionalist approaches, especially the emergentist approach (Davis & Bedore, 2013), seem to be the only stream of phonological theory currently accounting for this (see Sections 1.1 and 1.2).

Second, children's performances revealed a great influence of the ambient language phonology as well as an interaction of child-internal mechanisms with environmental factors (e.g., quantity and quality of language input, language prestige) which implies that the data are only broadly generalisable to other bilingual populations. From the current state of knowledge, it seems unavoidable to study each language pair separately (e.g., in order to identify CLT phenomena accurately) and to take into account as many environmental factors as possible in order to be able to determine the exact phonetic-phonological characteristics in a given bilingual population.

Third, the differences in the rate of acquisition between Turkish-German bilingual and monolingual children of each language highlight the inappropriateness of basing clinical decisions solely on monolingual norms (McLeod et al., 2017; Yavaş & Goldstein, 1998) as the bilinguals' broadly decelerated acquisition compared to monolinguals may be misinterpreted as a clinical delay.

Fourth, the qualitative differences regarding the types of phonological patterns occurring in monolinguals and bilinguals may lead to an over- identification of children with SSD since some phonological patterns Turkish-German bilingual children produced were unknown for or considered to be deviant for monolingual children (cf. Dodd et al., 1997). Hence, the assessment and evaluation of bilingual phonological skills is likely to be more time-consuming.

Fifth, the phonological development in bilinguals cannot generally be assumed to proceed linearly in both of their languages. It seems as if phases of re-organisations/updates of children's phonological systems, which appeared as stagnations or regressions at some time-points (Davis & Bedore, 2013; Werker et al., 2004), occur more frequently or persist longer in bilingual than in monolingual children (Kim et al., 2017). Hence, longitudinal investigations are particularly important for clinical settings to prevent misinterpretations as discussions in Section 9.2.1 have shown.

Sixth, based on the longitudinal investigation and categorisation at both time-points (see Chapter 9) the following actions for clinical decision-making can be derived:

- Children with a typical phonological performance in both languages are highly likely to perform typically at a later assessment as well. A follow-up of these children is rather optional.
- Children with atypical phonological performances in both languages are at risk for speech difficulties that persist over time and should therefore be (a) reassessed in regular intervals (e.g., 3-month intervals) to monitor their progress and (b) help them to receive phonological intervention to overcome their weaknesses.
- Children with atypical performances in only one of their languages should be monitored in 3 - 6 months' intervals as they are at risk for showing a delay/deviance in both of their languages at a later point in time. In addition, those children who continue to demonstrate atypical performances in the environmental language which are only slowly resolving may be considered for intervention if they are close to school enrolment (at the age of six in Germany) as the environmental language is going to be the language of instruction.
- Children's phonological abilities should be assessed in both of their languages at all time-points to ensure that both their weaker and stronger language are examined and evaluated in order to prevent misinterpretations.

10.3 Future directions

With the present research, a number of questions – especially all research questions – could be answered. However, the project also brought up further questions which require future research. The suggestions derived from this thesis are now briefly summarised.

Results of the categorisation process have revealed the importance of longitudinal data for monitoring children's phonological abilities but also to identify atypical phonological development (Kim et al., 2017). Since there was a large inter- and intra-individual variability in children's performances and normative data are very difficult to obtain from a population as heterogeneous as bilingual speakers, the question arises whether children's shape/course of phonological development (e.g., U-shape, fluctuating developmental trajectory) may be a more important marker for the identification of SSD than the deviation from normative scores gained with cross-sectional investigations as would be in line with Kim et al. (2017). In order to explore this aspect, more longitudinal studies on Turkish-German bilingual children's phonological development including more than two time-points would be required. Ideally this would include both typically and atypically developing children to explore a range of (hopefully) different developmental phonological trajectories.

A conduction of further longitudinal studies would additionally enable the verification of the clinical markers identified in this research. Especially, if psycholinguistic assessment tasks were included at every time-point – as opposed to only at the last assessment. This would, for instance, give insight into whether the identified markers are only valid for a specific age range.

Further, regression analyses of the present research did not reveal phone imitation skills to be a significant unique predictor of any phonological outcome measure. This may be the case since the relationship between phonetics and phonemics was explored. It would be interesting to see if phones that could be imitated correctly in isolation also occurred (as phones) in children's single-word productions or rather the other way around (i.e., whether sounds that could not be produced phonetically in the single-word productions could be imitated in isolation; see Section 9.2.2.3). Given the large unexplained variance, also further influential factors on children's phonological acquisition should be explored.

Moreover, it has frequently been hypothesised within this research that children's low item familiarity with the Turkish (but also some German) test items may have biased the phonological outcome (see Sections 5.1.2, 9.1, and 9.2.2). It would, therefore, be interesting to see to what extent children's performances would change if items with very low item familiarity scores were exchanged by other, better known words. Also, it would need to be explored whether this had an effect on children's categorisation or not. This would give insight into how “well-known” items need to be to yield to valid and reliable results within a phonology assessment. In this respect, it further needs to be noted that, at this point, native speakers of Turkish are required to administer the TPA. This limits its applicability to clinical practice since SLTs in Germany only rarely speak Turkish. It would therefore be necessary to explore further ways to administer the test, for example, in form of a digitalised version, so that many SLTs in Germany can benefit from it. A proposal for such a project has already been submitted to and funding has been approved by the European University of Applied Sciences.

Finally, it seems recommendable to include the assessment of children’s vocabulary knowledge (receptive and expressive) in future research, since this would allow for an estimation of the stability of their phonological representations (Macrae et al., 2014; Sosa, 2015; Stoel-Gammon, 2007; Vogel Sosa & Stoel-Gammon, 2006).

10.4 Summary and conclusion

Data of the present research have provided insight into important aspects of 3;0- to 6;5-year-old Turkish-German bilingual children’s phonological acquisition in both of their languages. In order to complete the picture and to further our understanding of children’s entire phonological acquisition process, it would be necessary to extent the analyses to phonetic aspects as well as vowels. Further, given that some of the 3;0-year-old children had already achieved a very high level of phonological competence (i.e., nearly completed consonant and CC inventories, no use of structural or systemic phonological patterns, low number of InfVar) and some 6;5-year-old children were still in the process of acquiring aspects of the phonological systems of their languages, it would be revealing to assess both younger and older Turkish-German bilingual children. First, this would provide insight into children’s very early acquisition behaviour. Second, it might shed light on the age of completion of their phonological acquisition. Lastly, the analysis of influential factors yielded some child-internal as well as -external factors that are most probably influencing Turkish-German bilingual children’s phonological acquisition. However, results also showed that these could not account for all of the variance in the data suggesting that further child-internal and/or environmental factors have an impact. An exploration of these would help to understand the interplay between child-internal and environmental factors in children’s speech acquisition, which is been very much put forward by representatives of functionalist approaches (see Section 1.1).

With regard to the differentiation of typical from atypical phonological acquisition, it can be concluded that the phonological and psycholinguistic markers identified seem to help with the clinical decision-making for children’s concurrent performance. The next steps would be to (a) verify the categorisation scheme and (b) examine its stability over time (i.e., also investigating the discriminative power of psycholinguistic tasks over time). Regarding (a), it would first of all be necessary to collect normative data regarding children's performances on the identified markers to facilitate the decision of what is actually typical. Second, it seems necessary to consider children's language experience patterns and examine whether these actually explain the category shifts from unable to be categorised to typical and reverse. Concerning (b), it first needs to be investigated whether the type of markers can also be used with younger and older children than the ones in the present research, since Stackhouse and colleagues (2007; 1997) claim that the level of difficulty in psycholinguistic tasks needs to be adjusted to children's age. Second, with regard to clinical practice it would be necessary to find out whether children actually stay in the categories they have been allocated to with the psycholinguistic markers. This is especially necessary since the number of unique contributions by the psycholinguistic markers to explaining phonological output was small. But still the weak performance in only one of them was used to allocate children to a certain category.

Overall, results suggest that this research project has laid a good foundation for the investigation of Turkish-German bilingual children's phonological acquisition and its differentiation from atypical development, that future research can build on.

REFERENCES

AERA, APA, & NCME. (1999). *Standards for educational and psychological testing*. Washington: American Educational Research Association.

Ambridge, B., & Lieven, E. V. M. (2013). *Child language acquisition. Contrasting theoretical approaches*. Cambridge: Cambridge University Press.

Anastasi, A. (1982). *Psychological testing* (5th ed.). New York: Macmillan Publishing.

Anderson, R. T. (2004). Phonological acquisition in preschoolers learning a second language via immersion: A longitudinal study. *Clinical Linguistics & Phonetics, 18*(3), 183-210.

Androutsopoulos, J. (2001). From the streets to the screens and back again: On the mediated diffusion of ethnolectal patterns in contemporary German. *LAUD Linguistic Agency, A522*, 1-24.

Armon-Lotem, S., de Jong, J., & Meir, N. (2015). *Assessing multilingual children: Disentangling bilingualism from language disorder*. Bristol: Multilingual Matters.

Babatsouli, E., & Ingram, D. (2015). What bilingualism tells us about phonological acquisition. In R. H. Bahr & E. R. Silliman (Eds.), *Routledge handbook of communication disorders* (pp. 173-182). New York: Routledge.

Backus, A. (2006). Turkish as an immigrant language in europe. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 689-724). Malden, MA: Blackwell.

Backus, A., Jørgensen, J. N., & Pfaff, C. (2010). Linguistic effects of immigration: Language choice, codeswitching, and change in western european Turkish. *Language and Linguistics Compass, 7*(4), 481–495.

Baker, E., Croot, K., McLeod, S., & Paul, R. (2001). Psyholinguistic models of speech development and their application to clinical practice. *Journal of Speech, Language, and Hearing Research, 44*(3), 685-702.

Ball, M. J., Müller, N., & Munro, S. (2006). Phonological development and disorder of bilingual children acquiring Welsh and English. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 346-382). Clevedon: Multilingual Matters.

BAMF - Bundesamt für Migration und Flüchtlinge (2016). *Das Bundesamt in Zahlen 2015. Asyl, Migration und Integration*. Retrieved 03/05/2017, from http://www.bamf.de/DE/Infothek/Statistiken/Asylzahlen/BundesamtInZahlen/bundesamt-in-zahlen-node.html.

Bankson, N. W., Bernthal, J. E., & Flipsen, P., Jr. (2013). Speech sound assessment procedures. In J. E. Bernthal, N. W. Bankson & P. Flipsen, Jr. (Eds.), *Articulation and phonological disorders. Speech sound disorders in children* (7. ed., pp. 180-211). New York: Pearson.

Barbour, S., & Stevenson, P. (1990). *Variation in German. A critical approach to German sociolinguistics*. Cambridge: Cambridge University Press.

Barlow, J. A., & Gierut, J. A. (1999). Optimality Theory in phonological acquisition. *Journal of Speech, Language, and Hearing Research, 42*(6), 1482-1498.

Battle, D. E. (2012). Communication disorders in a multicultural and global society. In D. E. Battle (Ed.), *Communication disorders in multicultural and international populations* (4th ed., pp. 2-19). St. Louis: Mosby.

Becker, B. (2010). Wer profitiert mehr vom Kindergarten? *Kölner Zeitschrift für Soziologie und Sozialpsychologie, 62*(1), 139-163.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*(1), 1-29. doi: 10.2167/beb392.0

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K., . . . Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish-English bilingual children. *Bilingualism: Language and Cognition, 15*(3), 616-629.

Ben-David, A., & Berman, R. A. (2007). Israeli Hebrew speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 437-456). New York: Thomson Delmar Learning.

Bernhardt, B. (1992). Developmental implications of nonlinear phonological theory. *Clinical Linguistics & Phonetics, 6*(4), 259-281.

Bernhardt, B. H., & Holdgrafer, G. (2001). Beyond the basics I: The need for strategic sampling for in-depth phonological analysis. *Language, Speech, and Hearing Services in Schools, 32*, 18-27.

Bernhardt, B. M., Romonath, R., & Stemberger, J. P. (2015). A comparison of fricative acquisition in German- and Canadian English-speaking children with protracted phonological development. In M. Yavas (Ed.), *Unusual productions in Phonology. Universal and language-specific considerations* (pp. 91-116). New York: Psychology Press.

Bernthal, J. E., Grossman, C. B., & Goll, A. E. A. (1989). Phonologically delayed children's responses to three types of pictured stimuli. *Journal of Childhood Communication Disorders, 12*(2), 137-143.

Bhatia, T. K., & Ritchie, W. C. (1999). The bilingual child: Some issues and perspectives. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of child language acquisition* (pp. 569-646). San Diego: Academic Press.

Bleile, K. M., & Tomblin, J. B. (1991). Regressions in the phonological development of two children. *Journal of Psycholinguistic Research, 20*(6), 483-499.

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. *Journal of Speech, Language & Hear Res, 58*(6), 1747-1760.

Boersma, P. (1998). *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives*. Den Haag: Hollan Academic Graphics/IFOTT.

Bosch, L., & Sebastian-Galles, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech, 46*(2-3), 217-243.

Bradford, A., & Dodd, B. (1996). Do all speech-disordered children have motor deficits? *Clinical Linguistics & Phonetics, 10*(2), 77-101.

Brandeker, M., & Thordardottir, E. (2015). Language exposure in bilingual toddlers: Performance on nonword repetition and lexical tasks. *American Journal of Speech-Language Pathology, 24*(2), 126-138.

Brendemoen, B. (1998). Turkish dialects. In L. Johanson & É. Á. Csató (Eds.), *The Turkic languages* (pp. 236-241). London: Routledge.

Brice, A. E., Carson, C. K., & O'Brien, J. D. (2009). Spanish-English articulation and phonology of 4- and 5-year-old preschool children: An initial investigation. *Communication Disorders Quarterly, 31*(1), 3-14.

Broomfield, J., & Dodd, B. (2004). The nature of referred subtypes of primary speech disability. *Child Language Teaching and Therapy, 20*(2), 135-151.

Browman, C. P., & Goldstein, L. (1992). Articulatory Phonology: An overview. *Phonetica, 49*, 155-180.

Bücker, T. (2007). Ethnolektale Varietäten des Deutschen im Sprachgebrauch Jugendlicher. *Studentische Arbeitspapiere zu Sprache und Interaktion*(9), 1-124.

Bunta, F., Fabiano-Smith, L., Goldstein, B. A., & Ingram, D. (2009). Phonological whole-word measures in 3-year-old bilingual children and their age-matched monolingual peers. *Clinical Linguistics & Phonetics, 23*(2), 156-175.

Burrows, L., & Goldstein, B. A. (2010). Whole word measures in bilingual children with speech sound disorders. *Clinical Linguistics & Phonetics, 24*(4-5), 357-368.

Büttner, J. (2012). *Phonologische Entwicklung im Deutschen bei bilingual russisch-deutsch aufwachsenden Kindern im Alter von 3;0 - 4;8 Jahren*. Unveröffentlichte Bacherlorarbeit. Hochschule Fresenius. Hamburg.

Bybee, J. L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Byers-Heinlein, K., Burns, T. C., & Werker, J. F. (2010). The roots of bilingualism in newborns. *Psychological Science, 21*(3), 343-348.

Caesar, L. G., & Kohler, P. D. (2007). The state of school-based bilingual assessment: Actual practice versus recommended guidelines. *Language, Speech, and Hearing Services in Schools, 38*(3), 190-200.

Campbell, T. F., Dollaghan, C. A., Rockette, H. E., Paradise, J. L., Feldmann, H. M., Shriberg, L. D., . . . Kurs-Lasky, M. (2003). Risk factors for speech delay of unknown origin in 3-year-old children. *Child Development, 74*(2), 346–357.

Cheung, P., & Abberton, D. (2000). Patterns of phonological disability in Cantonese-speaking children. *International Journal of Language and Communication Disorders, 35*, 451-474.

Chiat, S. (2015). Nonword repetition. In S. Armon-Lotem, J. de Jong & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language disorder* (pp. 125-147). Bristol: Multilingual Matters.

Chilla, S., Rothweiler, M., & Babur, E. (2010). *Kindliche Mehrsprachigkeit. Grundlagen - Störungen - Diagnostik*. München: Ernst Reinhardt.

Chlosta, C., Ostermann, T., & Schroeder, C. (2003). Die „Durchschnittsschule“ und ihre Sprachen: Ergebnisse des Projekts Sprachenerhebung Essener Grundschulen (SPREEG). *Essener Linguistische Skripte - elektronisch, 3*(1), 43-139.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Clausen, M. C., & Fox-Boyer, A. V. (submitted). Differential diagnosis of speech sound disorders in Danish-speaking children.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. E. Beckman (Eds.), *Papers in labortory phonology* (pp. 283-333). Cambridge: Cambridge University Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.

Cooperson, S. J., Bedore, L. M., & Peña, E. D. (2013). The relationship of phonological skills to language skills in Spanish-English-speaking bilingual children. *Clinical Linguistics & Phonetics, 27*(5), 371-389.

Core, C., & Scarpelli, C. (2015). Phonological development in young bilinguals: Clinical implications. *Seminars in Speech and Language, 36*(2), 100-108.

Crosbie, S., Holm, A., & Dodd, B. (2005). Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language & Communication Disorders, 40*(4), 467-491.

Cruz-Ferreira, M. (2012). Sociolinguistic and cultural considerations when working with multilingual children. In S. McLeod & B. A. Goldstein (Eds.), *Mutlilingual aspects of speech sound disorders in children* (pp. 13-23). Bristol: Mutlilingual Matters.

Crystal, D. (2010). *The Cambridge encyclopedia of language* (3rd ed.). Cambridge: Cambridge University Press.

Darcy, I., & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics, 40*(4), 568-581.

Davis, B. L., & Bedore, L. M. (2013). *An emergence approach to speech acquisition. Doing and knowing*. East Sussex: Psychology Press.

De Castro, M. M., & Wertzner, H. F. (2012). Stimulability: Auxiliary measure in the identifcation of difficulty in speech sounds production. *Jornal da Sociedade Brasileira de Fonoaudiologia, 24*(1), 49-56.

De Houwer, A. (2007). Parental language input patterns and children's bilingual use. *Applied Psycholinguistics, 28*, 411-424.

De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol: Multilingual Matters.

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders, 46*(6), 613-627.

Diggle, P. J., Heagerty, P., Liang, K.-Y., & Zeger, S. L. (2002). *Analysis of longitudinal data* (2nd ed.). Oxford: Oxford University Press.

Dixon, L. Q., Wu, S., & Daraghmeh, A. (2012). Profiles in bilingualism: Factors influencing kindergartners' language proficiency. *Early Childhood Education Journal, 40*(1), 25-34.

Dodd, B. (1995). *Differential diagnosis and treatment of children with speech disorder*. London: Whurr.

Dodd, B. (2005). *Differential diagnosis and treatment of children with speech disorder* (2nd ed.). London: Whurr.

Dodd, B. (2011). Differentiating speech delay from disorder: Does it matter? *Topics in Language Disorders, 31*(2), 96-111.

Dodd, B. (2014). Differential diagnosis of pediatric speech sound disorder. *Current Developmental Disorders Reports, 1*(3), 189-196.

Dodd, B., Holm, A., & Li, W. (1997). Speech disorder in preschool children exposed to Cantonese and English. *Clinical Linguistics & Phonetics, 11*(3), 229-243.

Dodd, B., Holm, A., Zhu, H., & Crosbie, S. (2003). Phonological development: A normative study of British English-speaking children. *Clinical Linguistics & Phonetics, 17*(8), 617-643.

Dodd, B., Leahy, J., & Hambly, G. (1989). Phonological disorders in children: Underlying cognitive deficits. *British Journal of Developmental Psychology, 7*(1), 55-71.

Dodd, B., So, L. K. H., & Li, W. (1996). Symptoms of disorder without impairment: The written and spoken errors of bilinguals. In B. Dodd, R. Campbell & L. Worrall (Eds.), *Evaluating theories of language: Evidence from disordered communication* (pp. 119-136). London: Whurr.

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic evaluation of articulation and phonology*. London: Psychological Corporation.

Donegan, P. (2013). Normal vowel development. In M. J. Ball & F. E. Gibbon (Eds.), *Handbook of vowels and vowel disorders* (pp. 24-60). New York; London: Psychology Press.

Donegan, P. J., & Stampe, D. (1979). The study of natural phonology. In D. A. Dinnsen (Ed.), *Current Approaches to Phonological Theory* (pp. 126-173). Bloomington & London: Indiana University Press.

dos Santos, C., & Ferré, S. (2016). A nonword repetition task to assess bilingual children's phonology. *Language Acquisition*, 1-14.

Duncan, T. S., & Paradis, J. (2016). English language learners' nonword repetition performance: The Influence of age, L2 vocabulary size, length of L2 exposure, and L1 phonology. *Journal of Speech, Language, and Hearing Research, 59*(1), 39-48.

Durrell, M. (1992). *Using German. A guide to contemporary usage*. Cambridge: Cambridge University Press.

Dyson, A. T., & Amayreh, M. M. (2007). Jordanian Arabic speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 288-299). New York: Thomson Delmar Learning.

Dyson, A. T., & Paden, E. P. (1983). Some phonological acquisition strategies used by two-year-olds. *Journal of Childhood Communication Disorders, 7*(1), 6-18.

Eadie, P., Morgan, A., Ukoumunne, O. C., Ttofari Eecen, K., Wake, M., & Reilly, S. (2014). Speech sound disorder at 4 years: Prevalence, comorbidities, and predictors in a community cohort of children. *Dev Med Child Neurol, 57*(6), 578-584.

Edwards, J., & Beckman, M. E. (2008). Methodological questions in studying consonant acquisition. *Clinical Linguistics & Phonetics, 22*(12), 937-956.

Ege, P. (2010). Türkçe'deki ünsüzlerin edinimi: Bir norm çalışması [Acquisition of consonants in Turkish: a normative study]. *Türk Psikoloji Dergisi, 25*(65), 16-39.

Eisenberg, S. L., & Hitchcock, E. R. (2010). Using standardized tests to inventory consonant and vowel production: A comparison of 11 tests of articulation and phonology. *Language, Speech, and Hearing Services in Schools, 41*(4), 488-503.

Elman, J. L. (2005). Connectionist models of cognitive development: Where next? *Trends in Cognitive Science, 9*(3), 111-117.

Evans, J. D. (1996). *Straightforward statistics for the behavioral sciences*. Pacific Grove: Brooks/Cole Publishing.

Fabiano-Smith, L., & Barlow, J. A. (2010). Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International Journal of Bilingual Education and Bilingualism, 13*(1), 81-97.

Fabiano-Smith, L., & Goldstein, B. A. (2010a). Early-, middle-, and late-developing sounds in monolingual and bilingual children: An exploratory investigation. *American Journal of Speech-Language Pathology, 19*(1), 66-77.

Fabiano-Smith, L., & Goldstein, B. A. (2010b). Phonological acquisition in bilingual Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research, 53*(1), 160-178.

Fabiano, L., & Goldstein, B. A. (2005). Phonological cross-linguistic effects in bilingual Spanish−English speaking children. *Journal of Multilingual Communication Disorders, 3*(1), 56-63.

Fachdienst Bürger, S. u. O. S. (2016, February). *Remscheider EinwohnerInnen am 31.12.2015*: Stadt Remscheid.

Fagan, S. M. B. (2009). *German. A linguistic introduction*. Cambridge: Cambridge University Press.

Fenson, L., Dale, P. S., Reznik, S., Bates, E., Thal, D. J., Pethick, S. J., . . . Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), i+iii-v+1-185.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*(2), 419-439.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage.

Flege, J. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effects of equivalence classification. *Journal of Phonetics, 15*, 47-65.

Flipsen, P., Jr., & Ogiela, D. A. (2015). Psychometric characteristics of single-word tests of children's speech sound production. *Language, Speech, and Hearing Services in Schools, 46*(2), 166-178.

Fox-Boyer, A., Casper, A., & Önal, N. (2014). *Speech disorders in bilingual children*. Paper presented at the 13th International Conference for the Study of Child Language, Amsterdam.

Fox-Boyer, A. V. (2014a). Ausspracheströungen im Deutschen. In A. Fox-Boyer (Ed.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Kindergartenphase* (pp. 41-54). München: Elsevier.

Fox-Boyer, A. V. (2014b). Phonologie-Erwerb. In A. Fox-Boyer (Ed.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Kindergartenphase* (pp. 9-15). München: Elsevier.

Fox-Boyer, A. V. (2014c). *PLAKSS II. Psycholinguistische Diagnostik kindlicher Aussprachestörungen II*. Frankfurt: Pearson Assessment.

Fox-Boyer, A. V., & Schäfer, B. (2015). Die phonetisch-phonologische Entwicklung von Kleinkindern (0-3 Jahre). In S. Sachse (Ed.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen. Kleinkindphase* (pp. 39-62). München: Elsevier.

Fox, A., & Brodbeck, D. (2004). Kinder mit Aussprachestörung ohne Behandlung. In A. Fox (Ed.), *Kindliche Aussprachestörungen. Phonologischer Erwerb, Diagostik und Therapie* (2nd ed., pp. 221-230). Idstein: Schultz-Kirchner.

Fox, A. V. (2000). *The acquisition of phonology and the classification of speech disorders in German-speaking children.* (Unveröffentlichte Dissertation), University of Newcastle Upon Tyne, Newcastle.

Fox, A. V. (2003). *Kindliche Aussprachestörungen. Phonologischer Erwerb – Differenzialdiagnostik – Therapie* (1. ed.). Idstein: Schulz-Kirchner.

Fox, A. V. (2004). *Kindliche Aussprachestörungen. Phonologischer Erwerb - Differenzialdiagnostik - Therapie* (2. ed.). Idstein: Schulz-Kirchner Verlag.

Fox, A. V. (2007). German speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 386-397). New York: Thomson Delmar Learning.

Fox, A. V. (2011). *Kindliche Aussprachestörungen. Phonologischer Erwerb - Differenzialdiagnostik - Therapie* (6. ed.). Idstein: Schulz-Kirchner Verlag.

Fox, A. V. (2016). *Kindliche Aussprachestörungen. Phonologischer Erwerb – Differenzialdiagnostik – Therapie* (7. ed.). Idstein: Schulz-Kirchner.

Fox, A. V., & Dodd, B. (1999). Der Erwerb des phonologischen Systems in der deutschen Sprache. *Sprache Stimme Gehör, 23*, 183-191.

Fox, A. V., & Dodd, B. (2001). Phonologically disordered German-speaking children. *American Journal of Speech-Language Pathology, 10*(3), 291-307.

Fox, A. V., Dodd, B., & Howard, D. (2002). Risk factors for speech disorders in children. *International Journal of Communication Disorders, 37*(2), 117-131.

Fredman, M. (2011). *Recommendations for working with bilingual children*. Retrieved 05/06/2015, from http://www.linguistics.ualberta.ca/en/CHESL_Centre/~/media/linguistics/Media/CHESL/Documents/WorkingWithBilingualChildren-May2011.pdf.

Friberg, J. C. (2010). Considerations for test selection: How do validity and reliability impact diagnostic decisions? *Child Language Teaching and Therapy, 26*(1), 77-92.

Fricke, S., & Schäfer, B. (2008). *Test für Phonologische Bewusstheitsfähigkeiten*. Idstein: Schultz-Kirchner.

Gagarina, N. (2014a). Die Erstsprache bei Mehrsprachigen im Migrationskontext. In S. Chilla & S. Haberzettl (Eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen* (Vol. Mehrsprachigkeit, pp. 19-37). München: Elsevier.

Gagarina, N. (2014b). Diganostik von Erstsprachkompetenzen im Migrationskontext. In S. Chilla & S. Haberzettl (Eds.), *Handbuch Spracherwerb und Sprachetnwicklungsstörungen* (Vol. Mehrsprachigkeit, pp. 73-84). München: Elsevier.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge? It all depends on the nonwords. *Memory & Cognition, 23*(1), 83-94.

Gathercole, V. C. M. (2014). Bilingualism matters: One size does not fit all. *International Journal of Behavioral Development, 38*(4), 359-366.

Gathercole, V. C. M., & Hoff, E. (2007). Input and the acquisition of language: Three questions. In E. Hoff & M. Shatz (Eds.), *Blackwell handbook of language development* (pp. 107-127). Oxford: Blackwell.

Gildersleeve-Neumann, C. E., & Goldstein, B. A. (2014). Cross-linguistic generalization in the treatment of two sequential Spanish-English bilingual children with speech sound disorders. *International Journal of Speech-Language Pathology*, 1-15.

Gildersleeve-Neumann, C. E., Kester, E. S., Davis, B. L., & Peña, E. D. (2008). English speech sound development in preschool-aged children from bilingual English-Spanish environments. *Language, Speech, and Hearing Services in Schools, 39*(3), 314-328.

Gildersleeve-Neumann, C. E., Peña, E. D., Davis, B. L., & Kester, E. S. (2009). Effects on L1 during early acquisition of L2: Speech changes in Spanish at first English contact. *Bilingualism: Language and Cognition, 12*(02), 259-272.

Gildersleeve-Neumann, C. E., & Wright, K. L. (2010). English speech acquisition in 3- to 5-year-old children learning Russian and English. *Language, Speech, and Hearing Services in Schools, 41*(4), 429-442.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar*. Oxon: Routledge.

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology*. Oxford: Basil Blackwell.

Goldstein, B. A., & Bunta, F. (2012). Positive and negative transfer in the phonological systems of bilingual speakers. *International Journal of Bilingualism, 16*(4), 388-401.

Goldstein, B. A., Bunta, F., Lange, J., Rodriguez, J., & Burrows, L. (2010). The effects of measures of language experience and language ability on segmental accuracy in bilingual children. *American Journal of Speech-Language Pathology, 19*(3), 238-247.

Goldstein, B. A., & Fabiano, L. (2007, February 13). Assessment and intervention for bilingual children with phonological disorders. *The ASHA Leader*.

Goldstein, B. A., Fabiano, L., & Iglesias, A. (2004). Spontaneous and imitated productions in Spanish-speaking children with phonological disorders. *Language, Speech, and Hearing Services in Schools, 35*(1), 5-15.

Goldstein, B. A., Fabiano, L., & Washington, P. S. (2005). Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools, 36*(3), 201-218.

Goldstein, B. A., & McLeod, S. (2012). Typical and atypical multilingual speech acquisition. In S. McLeod & B. A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 84-100). Bristol: Multilingual Matters.

Goldstein, B. A., & Pollock, K. (2004). Vowel production in Spanish-speaking children with phonological disorders: Dialect and sampling issues. *Journal of Multilingual Communication Disorders, 2*(2), 147-160.

Goldstein, B. A., & Washington, P. S. (2001). An initial investigation of phonological patterns in typically developing 4-year-old Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools, 32*(3), 153-164.

Goltz, R. H., & Walker, A. G. H. (1990). North Saxon. In C. V. J. Russ (Ed.), *The dialects of modern German* (pp. 31-58). London: Routledge.

Grassegger, H. (2004). *Phonetik / Phonologie* (2 ed.). Idstein: Schulz-Kirchner.

Grech, H., & Dodd, B. (2008). Phonological acquisition in Malta: A bilingual language learning context. *International Journal of Bilingualism, 12*(3), 155-171.

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain Lang, 36*(1), 3-15.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism, 1*(2), 131-149.

Grosjean, F. (2013). Bilingualism: A short introduction. In F. Grosjean & P. Li (Eds.), *The psycholinguistics of bilingualism* (pp. 5-21). Chichester: Wiley-Blackwell.

Gruber, F. A. (1999). Probability estimates and paths to consonant normalization in children with speech delay. *Journal of Speech, Language, and Hearing Research, 42*(2), 448-459.

Grunwell, P. (1985). Comment on the terms "phonetics" and "phonology" as applied in the investigation of speech disorders. *British Journal of Disorders of Communication, 20*(2), 165-170.

Grunwell, P. (1987). *Clinical phonology* (2. ed.). London: Croom Helm.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*, 267-288.

Hack, J., Marinova-Todd, S. H., & Bernhardt, M. B. (2012). Speech assessment of Chinese–English bilingual children: Accent versus developmental level. *International Journal of Speech-Language Pathology, 14*(6), 509-519.

Haig, G., & Braun, F. (1999). The state of the Turkish language in Germany. *TASG News. Newsletter of the Turkish Area Study Group, 49*, 13-18.

Hall, T. A. (2011). *Phonologie. Eine Einführung* (2nd ed.). Berlin: De Gruyter.

Hambly, H., Wren, Y., McLeod, S., & Roulstone, S. (2013). The influence of bilingualism on speech production: A systematic review. *International Journal of Language & Communication Disorders, 48*(1), 1-24.

Hammer, C. S., Lawrence, F., Rodriguez, B., Davison, M. D., & Miccio, A. W. (2011). Changes in language usage of Puerto Rican mothers and their children: Do gender and timing of exposure to English matter? *Applied Psycholinguistics, 32*(2), 275-297.

Hammer, C. S., & Rodriguez, B. (2012). Bilingual language acquisition and the child socialization process. In B. A. Goldstein (Ed.), *Bilingual language development* *and disorders in Spanish-English speakers* (2nd ed., pp. 31-46). Baltimore: Paul H. Brookes.

Hawkins, J. A. (2011). German. In B. Comrie (Ed.), *The World's Major Languages* (2 ed., pp. 86-109). Oxon: Routledge.

Hewlett, N. (1985). Phonological versus phonetic disorders: Some suggested modifications to the current use of the distinction. *British Journal of Disorders of Communication, 20*, 155-164.

Holm, A. (1998). *Speech development and disorder in bilingual children.* Newcastle University Upon Tyne, Newcastle.

Holm, A. (2007). Cantonese-influenced English speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 267-276). New York: Thomson Delmar Learning.

Holm, A., Crosbie, S., & Dodd, B. (2007). Differentiating normal variability from inconsistency in children's speech: Normative data. *International Journal of Language & Communication Disorders, 42*(4), 467-486.

Holm, A., & Dodd, B. (1999a). Differential diagnosis of phonological disorder in two bilingual children acquiring Italian and English. *Clinical Linguistics & Phonetics, 13*(2), 113-129.

Holm, A., & Dodd, B. (1999b). An intervention case study of a bilingual child with phonological disorder. *Child Language Teaching and Therapy, 15*(2), 139-158.

Holm, A., & Dodd, B. (1999c). A longitudinal study of the phonological development of two Cantonese-English bilingual children. *Applied Psycholinguistics, 20*(3), 349-376.

Holm, A., & Dodd, B. (2001). Comparison of cross-language generalisation following speech therapy. *Folia Phoniatrica et Logopaedica, 53*(3), 166-172.

Holm, A., & Dodd, B. (2006). Phonological development and disorder of bilingual children acquiring Cantonese and English. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 286-325). Clevedon: Multilingual Matters.

Holm, A., Dodd, B., & Ozanne, A. (1997). Efficacy of intervention for a bilingual child making articulation and phonological errors. *International Journal of Bilingualism, 1*(1), 55-69.

Holm, A., Dodd, B., Stow, C., & Pert, S. (1999). Identification and differential diagnosis of phonological disorder in bilingual children. *Language Testing, 16*(3), 271-292.

Hui, L. W., Wells, B., & Howard, S. (2015). Rate of multilingual phonological acquisition: Evidence from a cross-sectional study of English-Mandarin-Malay. *Clinical Linguistics & Phonetics, 29*(11), 793-811.

IEPMCS - International Expert Panel on Multilingual Children's Speech (2012). *Multilingual children with speech sound disorders: Position paper*. Retrieved 25/04/2013, from http://www.csu.edu.au/research/multilingual-speech/position-paper.

Iglesias, A., & Rojas, R. (2012). Bilingual language development of English language learners: Modeling the growth of two languages. In B. A. Goldstein (Ed.), *Bilingual language development and disorders in Spanish-English speakers* (2nd ed., pp. 3-30). Baltimore: Paul H. Brookes.

Ingram, D. (1989a). *First language acquisition: Method, description and explanation*. Cambridge: Cambridge University Press.

Ingram, D. (1989b). *Phonological disability in children* (2nd ed.). London: Whurr.

Isaac, K. M. (2007). Working with interpreters. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 122-126). NewYork: Thomson Delmar Learning.

Iuzzini, J., & Forrest, K. (2011). *Lexical and segmental inconsistency in disordered and typical developing speech*. Paper presented at the Annual convention of the American Speech-Language-Hearing Association, San Diego, CA.

iZotope. (2014). RX4 Audio Editor. Cambridge, Massachusetts: iZotope, Inc.

Jakobson, R. (1968). *Child language, aphasia and phonological universals*. Mouton: The Hauge.

James, D. G. (2001a). Use of phonological processes in Australian children aged 2 to 7;11 years. *Advances in Speech-Language Pathology, 3*(2), 109-127.

James, D. G., Ferguson, W. A., & Butcher, A. (2016). Assessing children's speech using picture-naming: The influence of differing phonological variables on some speech outcomes. *International Journal of Speech-Language Pathology, 18*(4), 364-377.

James, D. G., van Doorn, J., McLeod, S., & Esterman, A. (2008a). Patterns of consonant deletion in typically developing children aged 3 to 7 years. *International Journal of Speech-Language Pathology, 10*(3), 179-192.

James, D. G. H. (2001b). An item analysis of Australian English words for an articulation and phonological test for children aged 2 to 7 years. *Clinical Linguistics & Phonetics, 15*(6), 457-485.

James, D. G. H., van Doorn, J., & McLeod, S. (2008b). The contribution of polysyllabic words in clinical decision making about children's speech. *Clinical Linguistics & Phonetics, 22*(4-5), 345-353.

Johnson, C. E., & Lancaster, P. (1998). The development of more than one phonology: A case study of a Norwegian-English bilingual child. *International Journal of Bilingualism, 2*(3), 265-300.

Jordaan, H. (2008). Clinical Intervention for Bilingual Children: An International Survey. *Folia Phoniatrica et Logopaedica, 60*, 97-105.

Kalmár, M. (2006). Theorie und Praxis des Wiener Lautprüfverfahrens für Türkisch sprechende Kinder (WIELAU-T). *LogoTHEMA, 2*, 3-8.

Kehoe, M. (2011). Relationships between lexical and phonological development: A look at bilingual children - A commentary on Stoel-Gammon's 'Relationships between lexical and phonological development in young children'. *Journal of Child Language, 38*(1), 75-81.

Keim, I. (2007). Socio-cultural identity, communicative style, and their change over time: A case study of a group of German-Turkish girls in Mannheim/Germany. In P. Auer (Ed.), *Style and social identities. Alternative Approaches to Linguistic Heterogeneity* (pp. 155-186). Berlin: Walter de Gruyter.

Keller, R. E. (1961). *German dialects*. Manchester: Manchester University Press.

Kent, R. D. (1984). Psychobiology of speech development: Coemergence of language and a movement system. *American Journal of Physiology, 246*(Regulatory Integrative Comparative Physiology 15), R888-R894.

Keshavarz, H. M., & Ingram, D. (2002). The early phonological development of a Farsi-English bilingual child. *International Journal of Bilingualism, 6*(3), 255-269.

Khattab, G. (2006). Phonological acquisition by Arabic-English bilingual children. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 383-412). Clevedon: Multilingual Matters.

Khattab, G. (2007). Lebanese Arabic speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 300-312). New York: Thomson Delmar Learning.

Kim, J.-H., Ballard, E., & McCann, C. (2017). Age-related changes in segmental accuracy and error production in Korean-English bilingual children: Implications for clinical speech sound assessment procedures. *Speech, Language and Hearing*, 1-14.

Kim, J.-H., Ballard, E., & McCann, C. M. (2016). Error analysis of Korean-English bilingual children's speech productions. *Clinical Archives of Communication Disorders, 1*(1), 11-29.

Kirk, C., & Demuth, K. (2006). Accounting for variability in 2-year-olds' production of coda consonants. *Language Learning and Development, 2*(2), 97-118.

Kirk, C., & Vigeland, L. (2014). A psychometric review of norm-referenced tests used to assess phonological error patterns. *Language, Speech, and Hearing Services in Schools, 45*(4), 365-377.

Kirk, C., & Vigeland, L. (2015). Content coverage of single-word tests used to assess common phonological error patterns. *Language, Speech, and Hearing Services in Schools, 46*(1), 14-29.

Klein, H. B., & Liu-Shea, M. (2009). Between-word simplification patterns in the continuous speech of children with speech sound disorders. *Language, Speech, and Hearing Services in Schools, 40*(1), 17-30.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders, 43*(6), 456-473.

Kohnert, K. (2013). *Language disorders in bilingual children and adults*. San Diego: Plural Publishing.

Kopkallı-Yavuz, H. (2010). The sound inventory of Turkish: Consonants and vowels. In S. Topbaş & M. Yavaş (Eds.), *Communication disorders in Turkish* (pp. 27-47). Bristol: Mutlilingual Matters.

KopkallI-Yavuz, H., & Topbas, S. (2000). *Children's preferences in early phonological acquisition: How does it reflect sensitivity to the ambient language?* Proceedings of the Ninth International Conference on Turkish Linguistics, Oxford.

Kornfilt, J. (2011). Turkish and the Turcic languages. In B. Comrie (Ed.), *The world's major languages* (2nd ed., pp. 519-544). London: Routledge.

Kritikos, E. P. (2003). Speech-language pathologists' beliefs about language assessment of bilingual/bicultural individuals. *American Journal of Speech-Language Pathology, 12*, 73-91.

Kubaschk, L., Fox-Boyer, A. V., & Klann, J. (2015). Die Stimulierbarkeit isolierter Phone im Deutschen bei 2;5 – 4;0-Jährigen. *Sprache-Stimme-Gehör, 39*(Supplement), S5-S6.

Küppers, A., Şimşek, Y., & Schroeder, C. (2015). Turkish as a minority language in Germany: Aspects of language development and language instruction. *Zeitschrift für Fremdsprachenforschung, 26*(1), 29-51.

Lalkhen, A. G., & McCluskey, A. (2008). Clinical tests: Sensitivity and specificity. *Continuing Education in Anaesthesia, Critical Care & Pain, 8*(6), 221-223. doi: 10.1093/bjaceaccp/mkn041

Lammer, V., & Kalmár, M. (2004). *Wiener Lautprüfverfahren für Türkisch sprechende Kinder (WIELAU-T)*. Wien: Lernen mit Pfiff.

Langdon, H. W., & Quintanar-Sarellana, R. (2003). Roles and responsibilities of the interpreter in interactions with speech-language pathologists, parents, and students. *Seminars in Speech and Language, 24*(3), 235-244.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a systematic review of the literature. *International Journal of Language and Communication Disorders, 35*(2), 165-188.

Law, N. C. W., & So, L. K. H. (2006). The relationship of phonological development and language dominance in bilingual Cantonese-Putonghua children. *International Journal of Bilingualism, 10*(4), 405-427.

Lee, H. J., Kim, Y. T., & Yim, D. (2013). Non-word repetition performance in Korean-English bilingual children. *International Journal of Speech-Language Pathology, 15*(4), 375-382.

Lee, T., Ballard, E., & Purdy, S. (2015a). *Exploring diagnostic markers of speech sound disorders in Mandarin-English bilingual children in NZ*. Paper presented at the International Symposium on Monolingual and Bilingual Speech, Chania.

Lee, T., Ballard, E., & Purdy, S. (2015b). *Phonological development of successive Mandarin-English bilingual children in NZ*. Paper presented at the International Symposium on Monolingual and Bilingual Speech, Chania.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000). Academic outcomes in children with histories of speech sound disorders. *Journal of Communicaiton Discorders, 33*, 11-30.

Lewis, G. L. (1967). *Turkish Grammar*. Oxford: Oxford University Press.

Lidz, C. S., & Peña, E. D. (1996). Dynamic assessment. *Language Speech and Hearing Services in Schools, 27*(4), 367-372.

Lin, L. C., & Johnson, C. J. (2010). Phonological patterns in Mandarin-English bilingual children. *Clinical Linguistics & Phonetics, 24*(4-5), 369-386.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403-439). Dordrecht: Kluwer Academic Publishers.

Lindblom, B. (1992). Phonological units as adaptive emergents of lexical development. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 131-163). Timonium, MD: York Press.

Lindner, K., Mathieu, J., & Gagarina, N. (2016). *Verfahren zur Identifizierung bilingualer Kinder mit einem Risiko für USES: Sprachspezifisches Nachsprechen von Nicht-Wörtern*. Paper presented at the 9. Interdisziplinäre Tagung über Sprachentwicklungsstörungen (ISES IX), Heidelberg.

Lleó, C., & Kehoe, M. (2002). On the interaction of phonological systems in child bilingual acquisition. *International Journal of Bilingualism, 6*(3), 233-237.

Lleó, C., Kuchenbrandt, I., Kehoe, M., & Trujillo, C. (2003). Syllable final consonants in Spanish and German monolingual and bilingual acquisition. In N. Müller (Ed.), *(In)vulnerable Domains in Multilingualism* (pp. 191-220). Amsterdam: John Benjamins.

Lleó, C., & Prinz, M. (1996). Consonant clusters in child phonology and the directionality of syllable structure assignment*. *Journal of Child Language, 23*, 31-56.

Lleó, C., & Rakow, M. (2005). *Markedness effects in the acquisition of voiced stop spirantization by Spanish-German bilinguals.* Proceedings of the ISB4: Proceedings of the 4th International Symposium on Bilingualism, Somerville, MA.

Locke, J. L. (1983). *Phonological acquisition and change*. New York: Academic Press.

Lof, G. L. (1996). Factors associated with speech-sound stimulability. *Journal of Communication Disorders, 29*, 255-278.

Lorwatanapongsa, P., & Maroonroge, S. (2007). Thai speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 554-565). New York: Thomson Delmar Learning.

Lüke, C., & Ritterfeld, U. (2011). Mehrsprachige Kinder in sprachtherapeutischer Behandlung: Eine Bestandsaufnahme. *Heilpädagogische Forschung, 37*(4), 188-197.

Lydea, W. E. G., Brebner, C., & McCormack, P. (2014). A preliminary report on the English phonology of typically developing English-Mandarin bilingual preschool Singaporean children. *International Journal of Language & Communication Disorders, 49*(3), 317-332.

Macken, M. A., & Ferguson, C. A. (1983). Cognitive aspects of phonological development. In K. E. Nelson (Ed.), *Children's Language* (Vol. 4, pp. 255-282). Hillsdale, NJ: Lawrence Erlbaum.

MacLeod, A. A. N. (2013). Understanding bilingual phonological development. In A. A. N. MacLeod & B. Peter (Eds.), *Comprehensive perspectives on speech sound development and disorders: Pathways from linguistic theory to clinical practice* (pp. 305-324). New York: Nova Science.

MacLeod, A. A. N., Laukys, K., & Rvachew, S. (2011). The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. *International Journal of Speech-Language Pathology, 13*(6), 490-499.

MacNeilage, P. F. (1998). The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences, 21*(4), 499-511.

Macrae, T., Tyler, A. A., & Lewis, K. E. (2014). Lexical and phonological variability in preschool children with speech sound disorder. *American Journal of Speech-Language Pathology, 23*(1), 27-35.

MacWhinney, B. (2005a). Extending the Competition Model. *International Journal of Bilingualism, 9*(1), 69– 84.

MacWhinney, B. (2005b). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49-67). Oxford: Oxford University Press.

MacWhinney, B. (2008). A unified model. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 341-371). New York/Oxon: Routledge.

Maneva, B., & Genesee, F. (2002). *Bilingual babbling: Evidence for language differentiation in dual language acquisition.* Proceedings of the 26th Boston University Conference on Language Development, Sommerville, MA.

Martin, D. (2009). *Language disabilities in cultural and linguistic diversity*. Bristol: Multilingual Matters.

Masterson, J. J., Bernhardt, B. H., & Hofheinz, M. K. (2005). A comparison of single words and conversational speech in phonological evaluation. *American Journal of Speech-Language Pathology, 14*(3), 229-241.

Mayr, R., Howells, G., & Lewis, R. (2015). Asymmetries in phonological development: The case of word-final cluster acquisition in Welsh-English bilingual children. *Journal of Child Language, 42*(1), 146-179.

McCarthy, J. J., & Prince, A. S. (1998). Prosodic morphology. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphology* (pp. 283-305). Oxford: Blackwell.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders, 49*, 34-42.

McCormack, J., McLeod, S., Harrison, L. J., & McAllister, L. (2010). The impact of speech impairment in early childhood: Investigating parents' and speech-language pathologists' perspectives using the ICF-CY. *Journal of Communication Disorders, 43*(5), 378-396.

McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology, 11*(2), 155-170.

McLeod, S. (2012a). Multilingual speech assessment. In S. McLeod & B. A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 113-143). Bristol: Multilingual Matters.

McLeod, S. (2012b). Translation to practice: Creating sampling tools to assess multilingual children's speech. In S. McLeod & B. A. Goldstein (Eds.), *Multilingual aspects of speech sound disorders in children* (pp. 144-153). Bristol: Multilingual Matters.

McLeod, S. (2013). Speech sound acquisition. In J. E. Bernthal, N. W. Bankson & P. Flipsen Jr. (Eds.), *Articulation and phonological disorders. Speech sound disorders in children* (7th ed., pp. 58-113). New York: Pearson.

McLeod, S. (Ed.). (2007). *The international guide to speech acquisition*. New York: Thomson Delmar Learning.

McLeod, S., & Baker, E. (2014). Speech-language pathologists' practices regarding assessment, analysis, target selection, intervention, and service delivery for children with speech sound disorders. *Clinical Linguistics & Phonetics, 28*(7-8), 508-531.

McLeod, S., & Harrison, L. J. (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech, Language, and Hearing Research, 52*, 1213-1229.

McLeod, S., & Hewett, S. R. (2008). Variability in the production of words containing consonant clusters by typical 2- and 3-year-old children. *Folia Phoniatrica et Logopaedica, 60*(4), 163-172.

McLeod, S., Verdon, S., & Bowen, C. (2013). International aspirations for speech-language pathologists' practice with multilingual children with speech sound disorders: Development of a position paper. *Journal of Communication Disorders, 46*(4), 375-387.

McLeod, S., Verdon, S., & IEPMCS (2017). Tutorial: Speech assessment for multilingual children who do not speak the same language(s) as the speech-language pathologist. *American Journal of Speech Language Pathology*.

McReynolds, L. V., & Elbert, M. (1981). Criteria for phonologial process analysis. *Journal of Speech and Hearing Disorders, 46*, 197-204.

Meisel, J. M. (2004). The bilingual child. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 91-113). Oxford: Blackwell.

Menn, L. (1981). Theories of phonological development. *Annals of the New York academy of Sciences, 379*(1), 130-137.

Menn, L., Schmidt, E., & Nicholas, B. (2013). Challenges to theories, charges to a model: The Linked-Attractor model of phonological development. In M. M. Vihman & T. Keren-Portnoy (Eds.), *The emergence of phonology. Whole-word approaches and cross-linguistic evidence* (pp. 460-502). Cambridge: Cambridge University Press.

Mennen, I., & Okalidou, A. (2007). Greek speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 389-411). New York: Thomson Delmar Learning.

Mennen, I., & Stansfield, J. (2006). Speech and language therapy service delivery for bilingual children: A survey of three cities in Great Britain. *International Journal of Language & Communication Disorders, 41*(6), 635-652.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012-1027.

Miccio, A. W., Elbert, M., & Forrest, K. (1999). The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies. *American Journal of Speech-Language Pathology, 8*, 347–363.

Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual. Individual differences for language production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 389-407). Oxford: Oxford University Press.

Montanari, S. (2011). Phonological differentiation before age two in a Tagalog–Spanish–English trilingual child. *International Journal of Multilingualism, 8*(1), 5-21.

Montrul, S. A. (2008). *Incomplete acquisitionin bilingualism: Re-exmining the age factor*. Amsterdam: John Benjamins.

Morrison, J. A., & Shriberg, L. D. (1992). Articulation testing versus conversational speech sampling. *Journal of Speech and Hearing Disorders, 35*, 259-273.

Morrow, A., Goldstein, B. A., Gilhool, A., & Paradis, J. (2014). Phonological skills in English language learners. *Language, Speech, and Hearing Services in Schools, 45*(1), 26-39.

Mowrer, O. H. (1952). Speech development in the young child: 1. The autism theory of speech development and some clinical applications. *Journal of Speech and Hearing Disorders, 17*, 263-268.

Munro, S., Ball, M. J., Müller, N., Duckworth, M., & Lyddy, F. (2005). Phonological acquisition in Welsh–English bilingual children. *Journal of Multilingual Communication Disorders, 3*(1), 24-49.

Munson, B., Edwards, J., & Beckman, M. E. (2005). Relationships between nonword repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language, and Hearing Research, 48*, 61-78.

Muysken, P. (2013). Two linguistic systems in contact: Grammar, phonology, and lexicon. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (2nd ed., pp. 193-215). Oxford: Wiley-Blackwell.

Naş, V. (2010). *Türkisch-Artikulations-Test (TAT)*. Heidelberg: Springer.

Naş, V. (2015). Bedeutung muttersprachlich normierter Artikulationstests für bilinguale Kinder. *Forum Logopädie, 29*(1), 12-16.

Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004a). The development of early literacy skills among children with speech difficulties: A test of the "critical age hypothesis". *Journal of Speech, Language, and Hearing Research, 47*(2), 377–391.

Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004b). Educational consequences of developmental speech disorder: Key Stage I - National

Curriculum assessment results in English and mathematics. *British Journal of Educational Psychology, 74*, 173-186.
Neumann, S., Rietz, C., & Stenneken, P. (2016). *The German adaptation of the Intelligibility in Context Scale (ICS-G) − Reliability and Validity Evidence*. Paper presented at the 16. ICPLA Conference, Halifax.
Newbold, E. J., Stackhouse, J., & Wells, B. W. (2013). Tracking change in children with severe and persisting speech difficulties. *Clinical Linguistics & Phonetics, 27*(6-7), 521-539.
Newton, G. (1990). Central Franconian. In C. V. J. Russ (Ed.), *The dialects of modern German* (pp. 136-209). London: Routledge.
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
Olmsted, D. (1966). A theory of the child's learning of phonology. *Language, 42*, 531-535.
Ota, M., & Ueda, I. (2007). Japanese speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 457-471). New York: Thomson Delmar Learning.
Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism, 5*(1), 19-38.
Paradis, J., & Genesee, F. (1996). Syntactic acquisition in bilingual children: Autonomous or interdependent? *Studies in Second Language Acquisition, 18*(1), 1-25.
Paradis, J., Genesee, F., & Crago, M. B. (2011). *Dual language development & disorders: A handbook on bilingualism & second language learning* (2nd ed.). Baltimore: Brookes.
Parra, M., Hoff, E., & Core, C. (2011). Relations among language exposure, phonological memory, and language development in Spanish-English bilingually developing 2-year-olds. *Journal of Experimental Child Psychology, 108*(1), 113-125.
Pascoe, M., Stackhouse, J., & Wells, B. (2006). *Persisting speech difficulties in children*. Lodon: Wiley.
Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics, 28*, 399-410.
Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics, 18*(1), 41-58.
Pearson, B. Z., Fernandez, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning, 43*(1), 93-120.
Peña, E. D., Bedore, L. M., & Zlatic-Giunta, R. (2002). Category-generation performance of bilingual children: The influence of condition, category, and language. *Journal of Speech, Language & Hearing Research, 45*(3), 938-947.
Peña, E. D., Spaulding, T. J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology, 15*, 247-254.
Pfaff, C. (1993). Turkish language development in Germany. In G. Extra & L. Verhoeven (Eds.), *Immigrant languages in europe* (pp. 119-146). Clevedon: Multilingual Matters.
Pfaff, C. W. (1991). Turkish in contact with German: Language maintenance and loss among immigrant children in Berlin West. *International Journal of Sociology and Language, 90*, 97-129.
Phoon, H. S., Abdullah, A. C., & Maclagan, M. (2012). The effect of dialect on the phonological analysis of Chinese-influenced Malaysian English speaking children. *International Journal of Speech-Language Pathology, 14*(6), 487-498.
Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115-154.
Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development, 82*(6), 1834-1849.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*, 15-24.

Pons, F., Bosch, L., & Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychological Science, 26*(4), 490-498.

Powell, T. W., & Miccio, A. W. (1996). Stimulability: A useful clinical tool. *Journal of Communication Disorders, 29*(4), 237–253.

Preston, J. L., Hull, M., & Edwards, M. L. (2013). Preschool speech error patterns predict articulation and phonological awareness outcomes in children with histories of speech sound disorders. *American Journal of Speech-Language Pathology, 22*(2), 173-184.

Preston, J. L., & Seki, A. (2011). Identifying residual speech sound disorders in bilingual children: A Japanese-English case study. *American Journal of Speech-Language Pathology, 20*(2), 73-85.

Prezas, R. F., Hodson, B. W., & Schommer-Aikins, M. (2014). Phonological assessment and analysis of bilingual preschoolers' Spanish and English word productions. *American Journal of Speech-Language Pathology, 23*(2), 176-185.

Prince, A., & Smolensky, P. (2004). *Optimality theory: Constraint interaction in generative grammar*. Malden, MA: Blackwell Publishing.

Procter, A. L., Bunta, F., & Aghara, R. (2015). Stop VOT productions by young bilingual Spanish-English children and their monolingual peers. In M. Yavas (Ed.), *Unusual productions in phonology. Universals and language-specific considerations* (pp. 226-241). New York: Psychology Press.

Qualls, C. D. (2012). Multicultural and international research on communication disorders: Past, present, and future. In D. E. Battle (Ed.), *Communication* (pp. 296-310). St. Louis: Mosby.

Queen, R. (2001). Bilingual intonation patterns: Evidence of language change from Turkish-German bilingual children. *Language in Society, 30*, 55-80.

Rahilly. (2007). In S. McLeod (Ed.), *The international guide to speech acquisition*. New York: Thomson Delmar Learning.

Ramos, E., & Maed, J. (2014). Dual language intervention in a case of severe speech sound disorder. *Revista de investigación en Logopedia, 4*(2), 93-111.

Rathvon, N. (2004). *Early Reading Assessment: A practicioner's handbook*. New York: The Guilford Press.

Ray, J. (2002). Treating phonological disorders in a multilingual child: A case study. *American Journal of Speech-Language Pathology, 11*(3), 305-315.

RCSLT - Royal College of Speech and Language Therapists (2007). *Good practice for speech and language therapists. Working with clients from linguistic minority communities*. Retrieved 04/06/2015, from http://www.rcslt.org/members/publications/publications2/linguistic_minorities.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., . . . Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders, 49*(4), 416-451.

Reimers, P. M. (2015). Markedness in first language acquisition. In M. Yavaş (Ed.), *Unusual productions in phonology. Universals and language-specific considerations* (pp. 70-90). New York: Psychology Press.

Rinker, T., Budde-Spengler, N., & Sachse, S. (2016). The relationship between first language (L1) and second language (L2) lexical development in young Turkish-German children. *International Journal of Bilingual Education and Bilingualism, 20*(2), 218-233. doi: 10.1080/13670050.2016.1179260

Romaine, S. (1995). *Bilingualism* (2nd ed.). Oxford: Blackwell.

Roulstone, S., Loader, S., Northstone, K., & Beveridge, M. (2002). The speech and language of children aged 25 months: Descriptive data from the Avon Longitudinal Study of Parents and Children. *Early Child Development and Care, 172*(3), 259-268.

Roulstone, S., Peters, T. J., Glogowska, M., & Enderby, P. (2003). A 12-month follow-up of preschool children investigating the natural history of speech and language delay. *Child: Care, Health & Development, 29*(4), 245–255.

Ruiz-Felter, R., Cooperson, S. J., Bedore, L. M., & Peña, E. D. (2016). Influence of current input-output and age of first exposure on phonological acquisition in early bilingual Spanish-English-speaking kindergarteners. *International Journal of Language and Communication Disorders, 51*(4), 368-383.

Rust, J., & Golombok, S. (1999). *Modern psychometrics: The science of psychological assessment*. London: Routledge.

Rvachew, S., & Brosseau-Lapré, F. (2012). *Developmental phonological disorders: Foundations of clinical practice*. San Diego: Plural Publishing.

Saaristo-Helin, K. (2009). Measuring phonological development: A follow-up study of five children acquiring Finnish. *Language and Speech, 52*, 55-77.

Sachse, S., Budde-Spengler, N., & Rinker, T. (2016). *Türkischsprachiges Instrument zur frühen Wortschatzerfassung – Türkçe Ifade ve Lisan Dizelges Araştırması TILDA [Turkish language instrument for the early assessment of vocabulary skills].* University of Education Heidelberg/University of Ulm/University of Konstanz.

Salameh, E. K., Nettelbladt, U., & Norlin, K. (2003). Assessing phonologies in bilingual Swedish-Arabic children with and without language impairment. *Child Language Teaching and Therapy, 19*(3), 338-364.

Salgert, K. M. (2011). *Phonological acquisition in Turkish-German bilingual children*. Unpublished Master's dissertation. University of Sheffield. Sheffield.

Salgert, K. M., Fricke, S., & Wells, B. (in prep.). Phonological skills in young Turkish-German bilingual children.

Salgert, K. M., Fricke, S., & Wells, B. W. (2012). *Bilingualer Phonologieerwerb bei türkisch-deutsch-sprachigen Kindern*. Paper presented at the ISES VII, Leipzig.

Sánchez, L. (2006). Bilingualism/Second-language research and the assessment of oral proficiency in minority bilingual children. *Language Assessment Quarterly, 3*(2), 117-149.

Sankoff, G. (2004). Linguistic outcomes of language contact. In J. K. Chambers, P. Trudgill & N. Schilling-Estes (Eds.), *The handbook of language variation and change* (pp. 638-668). Malden, MA: Blackwell.

Savinainen-Makkonen, T. (2000). Word-initial consonant omissions - a developmental process in children learning Finnish. *First Language, 20*, 161-185.

Schaefer, B., & Fox-Boyer, A. V. (2016). The acquisition of initial consonant clusters in German-speaking 2-year olds. *International Journal of Speech-Language Pathology, Early online*, 1-14.

Schäfer, B., & Fox, A. V. (2006). Der Erwerb der Wortproduktionskonsequenz bei Zweijährigen: ein Mittel zur Früherkennung von Aussprachestörungen. *Sprache · Stimme · Gehör, 30*, 186-192.

Scharff Rethfeld, W. (2005). Das "Bilinguale Patientenprofil" als Basis einer logopädischen Intervention. *Forum Logopädie, 19*(3), 6-11.

Scharff Rethfeld, W. (2013). *Kindliche Mehrsprachigkeit. Grundlagen und Praxis der sprachtherapeutischen Intervention*. Stuttgart: Thieme.

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics, 31*(1), 117-140.

Schnitzer, M. L., & Krasinski, E. (1994). The development of segmental phonological production in a bilingual child. *Journal of Child Language, 21*(3), 585-622.

Schnitzer, M. L., & Krasinski, E. (1996). The development of segmental phonological production in a bilingual child: A contrasting second case. *Journal of Child Language, 23*(3), 547-571.

Schroeder, C., & Şimşek, Y. (2010). Die Entwicklung der Kategorie Wort im Türkisch-Deutsch bilingualen Schrifterwerb in Deutschland. In C. Weth (Ed.), *Schrifterwerb unter den Bedingungen von Mehrsprachigkeit und Fremdsprachenunterricht* (Vol. 37, pp. 55-79): Imis-Beiträge.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis, 25*(3), 263-286.

Sharp, K. M., & Mueller Gathercole, V. C. (2013). Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh-English bilingual children. *Child Language Teaching and Therapy, 29*(1), 77-89.

Shriberg, L. D. (1993). Four new speech and prosody-voice measures for genetics research and other studies in developmental phonological disorders. *Journal of Speech and Hearing Research, 36*, 105-140.

Shriberg, L. D. (1994). Five subtypes of developmental phonological disorders. *Clinics in Communication Disorders, 4*(1), 38-53.

Shriberg, L. D., & Austin, D. (1998). Comorbidity of speech-language disorder. Implications for a phenotype marker for speech delay. In R. Paul (Ed.), *Exploring the speech-language connection* (pp. 73-117). Baltimore: Paul H. Brookes.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997a). The percentage of consonants correct (PCC) metric: Extensions and reliability data. *Journal of Speech, Language, and Hearing Research, 40*(4), 708-722.

Shriberg, L. D., Austin, D., Lewis, B. A., McSweeny, J. L., & Wilson, D. L. (1997b). The Speech Disorders Classification System (SDCS): Extensions and lifespan and reference data. *Journal of Speech, Language, and Hearing Research, 40*(4), 723-740.

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lohmeier, H. L., McSweeny, J. L., . . . Wilson, D. L. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24*(10), 795-824.

Shriberg, L. D., Kent, R. D., Karlsson, H. B., McSweeny, J. L., Nadler, C. J., & Brown, R. L. (2003). A diagnostic marker for speech delay associated with otitis media with effusion: Backing of obstruents. *Clinical Linguistics & Phonetics, 17*(7), 529-547.

Shriberg, L. D., & Kwiatkowsky, J. (1982a). Phonological disorders I: A diagnostic classification system. *Journal of Speech and Hearing Disorders, 47*.

Shriberg, L. D., & Kwiatkowsky, J. (1982b). Phonological disorders III: A procedure for assessing severity in involvement. *Journal of Speech and Hearing Disorders, 47*, 256-270.

Shriberg, L. D., & Kwiatkowsky, J. (1994). Developmental phonological disorders: A clinical profile. *Journal of Speech and Hearing Research, 37*, 1100-1126.

Shriberg, L. D., Lewis, B. A., Tomblin, J. B., McSweeny, J. L., Karlsson, H. B., & Scheer, A. R. (2005). Toward diagnostic and phenotype markers for genetically transmitted speech delay. *Journal of Speech, Language, and Hearing Research, 48*, 834–852.

Shriberg, L. D., Lohmeier, H. L., Campbell, T. F., Dollaghan, C. A., Green, J. R., & Moore, C. A. (2009). A nonword repetition task for speakers with misarticulations: The syllable repetition task (SRT). *Journal of Speech, Language, and Hearing Research, 52*, 1189-1212.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year-old children and comorbidity with language impairment. *Journal of Speech, Language, and Hearing Research, 42*(6), 1461-1479.

Skahan, S. M., Watson, M., & Lof, G. L. (2007). Speech-language pathologists' assessment practices for children with suspected speech sound disorders: Results of a national survey. *American Journal of Speech-Language Pathology, 16*, 246–259.

Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century Crofts.

Smith, B. L., McGregor, K. K., & Demille, D. (2006). Phonological development in lexically precocious 2-year-olds. *Applied Psycholinguistics, 27*, 355–375.

Snowling, M. J., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics, 12*(3), 369-373.

So, L. K. H. (2007). Cantonese speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 313-326). New York: Thomson Delmar Learning.

So, L. K. H., & Dodd, B. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language, 22*(3), 473-495.

So, L. K. H., & Dodd, B. J. (1994). Phonologically disordered Cantonese-speaking children. *Clinical Linguistics & Phonetics, 8*(3), 235-255.

So, L. K. H., & Leung, C. S. S. (2006). Phonological development of Cantonese-Putonghua bilingual children. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 413-428). Clevedon: Multilingual Matters.

Sosa, A. V. (2015). Intraword variability in typical speech development. *American Journal of Speech-Language Pathology, 24*(1), 24-35.

Spaulding, T. J., Plante, E., & Farinella, K. A. (2006). Eligibility criteria for laguage impairment: Is the low end of normal always appropriate? *Language, Speech & Hearing Services in Schools, 37*(1), 61-72.

Stackhouse, J., Vance, M., Pascoe, M., & Wells, B. W. (2007). *Compendium of auditory and speech tasks. Children's speech and literacy difficulties* (Vol. Book 4). Chichester: John Wiley & Sons.

Stackhouse, J., & Wells, B. W. (1993). Psycholinguistic assessment in developmental speech disorders. *European Journal of Disorders of Communication, 28*, 331-348.

Stackhouse, J., & Wells, B. W. (1997). *Children's speech and literacy difficulties. A psycholinguistic model.* London: Whurr.

Stadt Essen, Amt für Statistik, Stadtforschung und Wahlen (2017). *Bevölkerung nach Geschlecht und Staatsangehörigkeit in den Stadtbezirken und Stadtteilen am 31.03.2017*. Retrieved 03/05/2017, from https://www.essen.de/rathaus/statistik/Statistik_Bevoelkerung.de.html.

Stadt Remscheid, der Oberbürgermeister (2016). *Statistisches Jahrbuch der Stadt Remscheid 2016.* Retrieved 03/05/2017, from http://www.remscheid.de/arbeiten-und-wirtschaft/standort-remscheid/daten-und-fakten/146380100000085019.php.

Stangen, I., Kupisch, T., Proietti Ergün, A. L., & Zielke, M. (2015). Foreign accent in heritage speakers of Turkish in Germany. In H. Peukert (Ed.), *Transfer effects in multilingual language development* (pp. 87-107). Amsterdam: John Benjamins.

Statistisches Amt für Hamburg und Schleswig-Holstein (2016). *Statistische Berichte. Bevölkerungsentwicklung in Hamburg 2015. Fortschreibung zum 31.12.2015 auf Basis des Zensus 2011*. Hamburg: Statistisches Amt für Hamburg und Schleswig-Holstein.

Statistisches Bundesamt (2017). *Bevölkerung und Erwerbstätigkeit. Bevölkerung mit Migrationshintergrund - Ergebnisse des Mikrozensus 2015*. Wiesbaden: Statistisches Bundesamt.

Stemberger, J. P., & Bernhardt, B. (1999). The emergence of faithfulness. In B. MacWhinney (Ed.), *The emergence of language* (pp. 417-446). Mahwah, NJ: Lawrence Earlbaum Associates.

Stoel-Gammon, C. (2007). Variability in speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 55-60). New York: Thomson Delmar Learning.

Stoel-Gammon, C. (2011). Relationships between lexical and phonological development in young children. *Journal of Child Language, 38*, 1-34.

Stoel-Gammon, C., & Dunn, C. (1985). *Normal and disordered phonology in children*. Baltimore: University Park Press.

Stoel-Gammon, C., & Williams, A. L. (2013). Early phonological development: Creating an assessment test. *Clinical Linguistics & Phonetics, 27*(4), 278-286.

Storkel, H. L. (2006). Do children still pick and choose? The relationship between phonological knowledge and lexical acquisition beyond 50 words. *Clinical Linguistics & Phonetics, 20*(7-8), 523-529.

Stow, C., & Pert, S. (1998). The development of a bilingual phonology assessment. *International Journal of Language & Communication Disorders, 33*(SUPPL. 1), 340-342.

Stow, C., & Pert, S. (2006). Phonological acquisition in bilingual Pakistani Heritage children in England. In H. Zhu & B. Dodd (Eds.), *Phonological development and* *disorders in children: A multilingual perspective* (pp. 326-345). Clevedon: Multilingual Matters.

Summers, C., Bohman, T. M., Gillam, R. B., Peña, E. D., & Bedore, L. M. (2010). Bilingual performance on nonword repetition in Spanish and English. *International Journal of Language & Communication Disorders, 45*(4), 480-493.

Tekin, Ö. (2011). Kontrastive Phonetik Deutsch-Türkisch. Unter besonderer Berücksichtigung der segmental-phonetischen und orthographischen Schwierigkeiten türkischer Deutschlerner. *IDV-Magazin, 84*, 601-615.

Tekin, Ö., & Colliander, P. (2010). Das Türkendeutsch. Phonetische Charakteristika und die Auswirkungen auf das Deutsche. *Zeitschrift für interkulturelle Germanistik, 1*(2), 49-61.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: The MIT Press.

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders, 46*(1), 1-16.

Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders, 4*(1), 1-21.

Topbaş, S. (1997). Phonological acquisition in Turkish children: Implications for phonological disorders. *European Journal of Disorders of Communication, 32*, 377-396.

Topbaş, S. (2006). Does the speech of Turkish-speaking phonologically disordered children differ from that of children speaking other languages? *Clinical Linguistics & Phonetics, 20*(7-8), 509-522.

Topbaş, S. (2007). Turkish speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 566-579). New York: Thomson Delmar.

Topbas, S., Kacar-Kutukcu, D., & Kopkalli-Yavuz, H. (2014). Performance of children on the Turkish Nonword Repetition Test: Effect of word similarity, word length, and scoring. *Clinical Linguistics & Phonetics, 28*(7-8), 602-616.

Topbaş, S., & KopkallI-Yavuz, H. (1997). *The onset of a linguistic system: Is there evidence from the acquisition of final devoicing in Turkish?* Proceedings of the GALA `97 conference on Language Acquisition, Edinburgh.

Topbaş, S., & Kopkalli-Yavuz, H. (2008). Reviewing sonority for word-final sonorant+obstruent consonant cluster development in Turkish. *Clinical Linguistics & Phonetics, 22*(10-11), 871-880.

Topbaş, S., & Ünal, Ö. (2010). An alternating treatment comparison of minimal and maximal opposition sound selection in Turkish phonological disorders. *Clinical Linguistics & Phonetics, 24*(8), 646-668.

Topbaş, S., & Yavaş, M. (2006). Phonological acquisition and disorders in Turkish. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 233-261). Clevedon: Multilingual Matters.

Trost, S. (2007-2014). *Wortgenerator. Buchstabenhäufigkeiten.* Retrieved 29/05/2014, from http://www.sttmedia.de/buchstabenhaeufigkeiten.

Tugay, E., & Schultz-Ünsal, F. (2013). Phonologische Prozesse bei türkisch und deutsch-türkisch sprechenden Kindern. *Forum Logopädie, 27*(3), 28-32.

Tyler, A. A., & Macrae, T. (2010). Stimulability: Relationships to other characteristics of children's phonological systems. *Clinical Linguistics & Phonetics, 24*(4-5), 300-310.

Ullrich, A. (2010). *Evidensbasierte Diagnostik phonologischer Störungen – Entwicklung und Evaluation eines Sprachanalyseverfahrens auf der Basis nichtlinearer phonologischer Theorien.* (PhD Unpublished PhD thesis), Universität zu Köln, Köln.

United Nations, Department of Economic & Social Affairs, Population Division (2016). *International migration report 2015 (ST/ESA/SER.A/384).* Retrieved 04/05/2017, from www.unpopulation.org.

Ünsal, F., & Fox, A. V. (2002). Lautspracherwerb bei zweisprachigen Migrantenkindern (Türkisch-Deutsch). *Forum Logopädie, 16*(3), 10-15.

van Riper, C. (1963). *Speech Correction: Principles and Methods* (4. ed.). New York: Englewood Cliffs, Prentice Hall.

Vance, M., Stackhouse, J., & Wells, B. (2005). Speech-production skills in children aged 3–7 years. *International Journal of Language & Communication Disorders, 40*(1), 29-48.

Vick, J. C., Campbell, T. F., Shriberg, L. D., Green, J. R., Truemper, K., Rusiewicz, H. L., & Moore, C. A. (2014). Data-driven subclassification of speech sound disorders in preschool children. *Journal of Speech, Language, and Hearing Research, 57*(6), 2033-2050.

Vihman, M. (2014). *Phonological development. The first two years* (2nd ed.). Oxford: Wiley Blackwell.

Vihman, M. (2016). Prosodic structures and templates in bilingual phonological development. *Bilingualism: Language and Cognition, 19*(1), 69-88.

Vihman, M. M. (1996). *Phonological development: The origins of language in the child*. Oxford: Basil Blackwell.

Vihman, M. M. (2002). Getting started without a system: From phonetics to phonology in bilingual development. *International Journal of Bilingualism, 6*(3), 239-254.

Vogel Sosa, A., & Bybee, J. L. (2011). A cognitive approach to clinical phonology. In M. J. Ball, M. R. Perkins, N. Müller & S. Howard (Eds.), *The handbook of clinical linguistics* (pp. 480-490). Oxford: Wiley-Blackwell.

Vogel Sosa, A., & Stoel-Gammon, C. (2006). Patterns of intra-word phonological variability during the second year of life. *Journal of Child Language, 33*, 31-50.

Wagner, L. (2008). *SCREEMIK 2. Screening der Erstsprachfähigkeit bei Migrantenkindern: Russisch-Deutsch, Türkisch-Deutsch*. München: Eugen Wagner Verlag.

Wagner, L. (2010). Screening der Erstsprachfähigkeit bei Migrantenkindern - SCREEMIK 2 als differenzialdiagnostisches Abgrenzungsmittel für russisch-deutsche und türkisch-deutsche Kinder. *Die Sprachheilarbeit, 55*(2), 54-64.

Waring, R., & Knight, R. (2013). How should children with speech sound disorders be classified? A review and critical evaluation of current classification systems. *International Journal of Language & Communication Disorders, 48*(1), 25-40.

Watson, M. M., & Scukanec, G. P. (1997). Profiling the phonological abilities of 2-year-olds: A longitudinal investigation. *Child Language Teaching and Therapy, 13*(1), 3-14.

Weigl, I., & Reddemann-Tschaikner, M. (2001). *HOT - ein handlungsorientierter Therapieansatz*. Stuttgart: Thieme.

Weiss, A. L. (2007). Personal factors and their influence on speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 91-95). New York: Thomson Delmar Learning.

Werker, J. F., Hall, D. G., & Fais, L. (2004). Reconstruing U-shaped functions. *Journal of Cognition and Development, 5*(1), 147-151.

Wiese, R. (2011). *Phonetik und Phonologie*. Paderborn: W. Fink.

Willard, J. A., Agache, A., Jäkel, J., Glück, C. W., & Leyendecker, B. (2014). Family factors predicting vocabulary in Turkish as a heritage language. *Applied Psycholinguistics, 36*(04), 875-898.

Williams, A. L., & Elbert, M. (2003). A prospective longitudinal study of phonological development in late talkers. *Language, Speech & Hearing Services in Schools, 34*(2), 138-153.

Williams, C. J., & McLeod, S. (2012). Speech-language pathologists' assessment and intervention practices with multilingual children. *International Journal of Speech-Language Pathology, 14*(3), 292-305.

Williams, N., & Chiat, S. (1993). Processing deficits in children with phonological disorder and delay: A comparison of responses to a series of output tasks. *Clinical Linguistics & Phonetics, 7*(2), 145-160.

Williams, P., & Stackhouse, J. (1998). Diadochokinetic skills: Normal and atypical performance in children agend 3-5 years. *International Journal of Language & Communication Disorders, 33*(Suppl), 481-486.

Williams, P., & Stackhouse, J. (2000). Rate, accuracy, and consistency: Diadochokinetic performance of young, normally developing children. *Clinical Linguistics & Phonetics, 14*(4), 267-293.

Winter, K. (2001). Numbers of bilingual children in speech and language therapy: Theory and practice of measuring their representation. *International Journal of Bilingualism, 5*(4), 465-495.

Wolk, L., & Meisler, A. W. (1998). Phonological assessment: A systematic comparison of conversation and picture naming. *Journal of Communication Disorders, 31*, 291-313.

Wren, Y. E., Roulstone, S. E., & Miller, L. L. (2012). Distinguishing groups of children with persistent speech disorder: Findings from a prospective population study. *Logopedics Phoniatrics Vocology, 37*(1), 1-10.

Wyatt, T. (2012). Assessment of multicultural and international clients with communication disorders. In D. E. Battle (Ed.), *Communication disorders in multicultural and international populations* (pp. 243-278). St. Louis: Mosby.

Yalcinkaya, F., Muluk, N. B., & Budak, B. (2010). Speech sounds acquisition evaluated by speech sound development test (SSDT) in Turkish-speaking children. *Journal of International Advanced Otology, 6*(1), 60-66.

Yang, H.-Y., & Zhu, H. (2010). The phonological development of a trilingual child: Facts and factors. *International Journal of Bilingualism, 14*(1), 105-126.

Yavaş, M. (1998). *Phonology development and disorders*. San Diego, London: Singular Publishing Group.

Yavaş, M. (2010). Acquisition of /s/-clusters in Spanish-English bilingual children with phonological disorders. *Clinical Linguistics & Phonetics, 24*(3), 188-198.

Yavaş, M. (Ed.). (2015). *Unusual productions in phonology. Universals and language-specific considerations*. New York: Psychology Press.

Yavaş, M., & Goldstein, B. A. (1998). Phonological assessment and treatment of bilingual speakers. *American Journal of Speech-Language Pathology, 7*(2), 49-60.

Yavaş, M., & Goldstein, B. A. (2006). Aspects of bilinugal phonology: The case of Spanish-English bilingual children. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 265-285). Clevedon: Multilingual Matters.

Yavaş, M., & Topbaş, S. (2004). Liquid development in Turkish: Salience vs. frequency. *Journal of Multilingual Communication Disorders, 2*(2), 110-123.

Yip, V. (2013). Simultaneous language acquisition. In F. Grosjean & P. Li (Eds.), *The psycholinguistics of bilingualism* (pp. 119-144). Chichester: Wiley-Blackwell.

Zajdó, K. (2007). Hungarian speech acquisition. In S. McLeod (Ed.), *The international guide to speech acquisition* (pp. 412-436). New York: Thomson Delmar Learning.

Zhu, H. (2000). *Phonological development and disorder of Putonghua (Modern Standard Chinese)-speaking children.* (PhD), Newcastle University upon Tyne, Newcastle.

Zhu, H., & Dodd, B. (2000a). Development and change in the phonology of Putonghua-speaking children with speech difficulties. *Clinical Linguistics & Phonetics, 14*(5), 351-368.

Zhu, H., & Dodd, B. (2000b). The phonological acquisition of Putonghua (Moden Standarad Chinese). *Journal of Child Language, 27*(1), 3-42.

Zhu, H., & Dodd, B. (2006). Towards developmental universals. In H. Zhu & B. Dodd (Eds.), *Phonological development and disorders in children: A multilingual perspective* (pp. 431-449). Clevedon: Multilingual Matters.

Ziller, D., & Wohlleben, B. (2006). Lautinventar und phonologische Prozesse bei 4-jährigen Kindern. *Die Sprachheilarbeit, 51*(2), 61-67.

Zimmer, K., & Orgun, O. (1999). Turkish. In T. I. P. Association (Ed.), *Handbook of the International Phonetic Association. A guide to the use of the International Phonetic Alphabet* (pp. 154-156). Cambridge: Cambridge University Press.

Zimmerer, F., Scharinger, M., & Reetz, H. (2011). When BEAT becomes HOUSE: Factors of word-final /t/-deletion in German. *Speech Communication, 53*(6), 941-954.

APPENDICES

Appendix A: Comparison of the psychometric criteria of the available phonology assessment tools for Turkish-German bilingual children

		Criteria	Assessments: WIELAU-T[1]	SCREEMIK-2[2]	TAT[3]
Linguistic criteria	Phonemes	All included?	No; /ʒ/ is missing	No; /ʒ, ɣ, h, j, t͡ʃ, d͡ʒ/ are missing	Yes
		All possible word positions?	Not for all consonants (9/21)	Not for all consonants (5/15)	Not for all consonants (14/21)
		≥ 4 - 5 occurrences per included consonant?	No; /g, h, s, j/: 3x /d͡ʒ/: 2x	No; /d, g/: 2x	No; /h, ʒ/: 3x /ɣ/: 2x
	Syllables	Test items' number of syllables?	1 - 4	1 – 3	1 - 4
		Consonant clusters?	Only /rt/ (1x)	No	Only /lp/ (1x)
	Phonological patterns	≥ 5 occurrences for all typical patterns?	No; CCR: 1x	No; CCR: 0x Deaf: 0x WFDev: 2x	No; CCR: 1x
		≥ 5 occurrences for all atypical patterns?	Yes	No[4]	Yes
	Vocabulary	Appropriate for target population?	Evidence N/A	Evidence N/A	Evidence N/A
	Item difficulty	Item familiarity analyses?	N/A	No	N/A
		Different syllable complexities?	Yes	Yes[5]	Yes

Note: to be continued on the next page

		Criteria	Assessments WIELAU-T[1]	SCREEMIK-2[2]	TAT[3]
Psychometric criteria	Validity	What type of analyses are conductible?	Segment-orientated descriptions of possible speech difficulties	Segment-orientated evaluation of children's realisations with *correct* / *incorrect* / *not appraisable*	Phonetic and phonemic inventory, phonetic-phonological pattern analysis*
		Cultural-specific words	No	Only one	Yes
		Cultural-sensitive words	Yes	Yes	Yes
		Cultural-specific pictures	No	Yes	Yes
		Cultural-sensitive pictures	Yes	Yes	Yes
	Reliability	Test-retest reliability	N/A	N/A	N/A
		Inter-rater reliability - Transcriptions - Classification	N/A	*N* = 15, Agreement of > 90 %[6] for analysis of entire screening (i.e., incl. vocabulary and grammar tasks)	N/A
		Intra-rater reliability - Transcriptions - Classification	N/A	N/A	N/A
		Internal (item) consistency	N/A	Cronbach's Alpha: > 0.82	N/A
	Norms	Normative investigation?	N/A	Yes (*N* = 388), incl. a subgroup of children with SLI	Yes (*N* = 136), but analysis procedures non-transparent

Note: continued from previous page
1Wiener Lautprüfverfahren für Türkisch sprechende Kinder (Lammer & Kalmár, 2004), **2**Screening der Erstsprachfähigkeit bei Migrantenkindern: Russisch-Deutsch, Türkisch-Deutsch (Wagner, 2008), **3**Türkisch-Artikulations-Test (Naş, 2010), **4**The pattern inconsistent use of /h/ is unlikely to occur since /h/ is not assessed in the screening. Further, it needs to be noted that the Screemik-II is designed to conduct a segment-based analysis so that phonological patterns could remain undetected if they do not affect a target phoneme in a target word, **5**Targeted phonemes are never assessed in syllable-final within word positions, **6**No precise percentages are mentioned in the handbook, neither the way they were calculated. **CCR**: Consonant cluster reduction, **Deaf**: Deaffrication, **WFDev**: Word-final Devoicing, *These analyses are recommended by the author but, as this table displays, the item set does not allow for them to be done thoroughly.

Appendix B: Ethical approval for t1

ETHICS REVIEWER'S COMMENTS FORM

This form is for use when ethically reviewing a research ethics application form.

1. Name of Ethics Reviewer:	Dr Richard Body Dr Sarah Spencer Dr Catherine Tattersall
2. Research Project Title:	Speech sound development in Turkish-German bilingual children
3. Principal Investigator (or Supervisor):	Dr Silke Fricke
4. Name of Student (if applicable):	Katharina Salgert
5. Academic Department / School:	Human Communication Sciences
6. I confirm that I do not have a conflict of interest with the project application	
7. I confirm that, in my judgment, the application should:	

Be approved:	**Be approved with *suggested* amendments in '8' below:** and/	**Be approved providing *requirements* specified in '9' below are met:**	**<u>NOT</u> be approved for the reason(s) given in '10' below:**
	X	X	

8. Approved with the following suggested, optional amendments (i.e. it is left to the discretion of the applicant whether or not to accept the amendments and, if accepted, the ethics reviewers do not need to see the amendments):
• Consider changing the wording of the questionnaire so that it is more accessible, e.g. some words like acquisition may be a barrier for some parents (consider changing to 'learning Turkish') • Some words such as 'media' could be more parent-friendly ('spare time'), • Examples may help parents provide accurate information, for example of grammatical mistakes • Consider changing 'generally typically developing' to more parent-friendly term. • Appendix 7: Consider changing 'thank you very much for allowing your child to take part' to 'agreeing for your child to take part', as this sounds more research-focused. • As you will be using letter-number codes – will you have 'linked anonymity' data? For example if a parent decided to withdraw after data collection would you be able to destroy their child's data? If not you need to make it clear on the information sheet that once data is anonymised you will not be able to destroy it as you would not be able to identify it. If you do have linked anonymity then you need to state that a child's data would be destroyed if they were withdrawn from the study.
9. Approved providing the following, compulsory requirements are met (i.e. the ethics reviewers <u>do not need</u> to see the required changes):
• Make sure that the potential suggestion of meeting with a paediatrician/SLT if assessment highlights delay is clear from the beginning on the project (on information leaflet and consent form).
10. Not approved for the following reason(s):
11. Date of Ethics Review: 18 June 2013

Appendix C: Ethical approval for t2

The University Of Sheffield.

Department Of Human Communication Sciences

Head of Department
Professor Patricia E. Cowell, BA, MS, PhD

362 Mushroom Lane
Sheffield
S10 2TS
United Kingdom

Telephone: +44 (0) 114 222 2418/ 2402/ 2405
International: +44 (0) 114 222 2418
Fax: +44 (0) 114 2222439
Email: hcs-support@sheffield.ac.uk
http://www.shef.ac.uk/hcs

To:
Dr Silke Fricke
Human Communication Sciences
University of Sheffield

08/09/2014

Dear Silke,

Re Project: Phonological acquisition and differential diagnosis of speech difficulties in Turkish-German bilingual children

I have taken Chair's Action on these amendements to your project which received ethics permission in June 2013 and can confirm that this proposed project can now go ahead.

Regards,

Professor Ray Wilkinson
Ethics Review Coordinator (Staff & PGR Projects)
Department of Human Communication Sciences

Appendix D: Parents' questionnaire at t1

Speech Sound Development in Turkish-German Bilingual Children

Parents' Questionnaire

As you agree for your child to take part in this research project you are kindly asked to provide some information about the bilingual situation in your family and your child's use of Turkish and German. This information will be treated as strictly confidential and will only be used to identify factors that may influence children's acquisition of speech sounds and their performance in the assessment.

As soon as we receive your completed questionnaire this first page containing your child's name will be destroyed and will be replaced by a research participant code on page 2. This questionnaire will be stored in a locked filing cabinet at the researcher's office together with the audio recordings. Only Katharina Salgert and her supervisors will be allowed access to this data which will be destroyed as soon as related publications are in press.

Your child's name: ______________________________

Child's Code (to be filled in by the researcher): ________________________________

01) What is your relation to the child? ☐ mother/female carer ☐ father/male carer

02) Mother tongue of the father/male carer: ______________________________

03) Mother tongue of the mother/female carer: ______________________________

Acquisition and Use of Turkish

A1) Who speaks Turkish to your child?
☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

A2) When did you or another family member start to speak Turkish to your child?
☐ from birth on when the child was: ☐ 1 year old ☐ 2 years old ☐ 3 years old
☐ 4 years old

A3) Please state which dialectal variations of Turkish your child hears on a regular basis?

____

A4) How regularly does your child **hear** Turkish?
☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

A5) When hearing Turkish, how many hours on average does your child **hear** Turkish per day? ☐

A6) My child speaks Turkish to:
☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

A7) How regularly does your child **speak** Turkish?
☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

A8) When speaking Turkish, how many hours on average does your child **speak** in Turkish per day? ☐

A9) Where does your child speak Turkish?
☐ at home ☐ in the nursery ☐ at friends' ☐ at relatives'
☐ at hobbies/leisure activities ☐ on holidays (in Turkey) ☐ in public

A10) How would you rate the proficiency of your child in Turkish? *(Please tick the box that applies the most.)*

☐ very good	☐ good	☐ average	☐ poor
understands (almost) everything, knows a large number of words, makes only age-appropriate grammatical mistakes	often understands what is being said, knows many words, makes minor grammatical mistakes	understands the general idea of what is being said, knows some words, makes grammatical mistakes	understands and speaks only a few words, does not speak full sentences

Acquisition and Use of German

B1) Who speaks German to your child?

☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

B2) When did your child start to hear German?

☐ from birth on when the child was: ☐ 1 year old ☐ 2 years old ☐ 3 years old
☐ 4 years old

B3) Does your child hear a different dialect al variation of German than the Hamburgian (Northern)?

☐ Yes If yes, please state which one: ______________________________
☐ No

B4) How regularly does your child **hear** German?

☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

B5) When hearing German, how many hours on average does your child **hear** German per day? ☐

B6) My child speaks German to:

☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

B7) How regularly does your child **speak** German?

☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

B8) When speaking German, how many hours does your child **speak** in German on average per day? ☐

B9) Where does your child speak German?

☐ at home ☐ in the nursery ☐ at friends' ☐ at relatives'
☐ at hobbies/leisure activities ☐ on holidays (in Turkey) ☐ in public

B10) How would you rate the proficiency of your child in <u>German</u>? *(Please tick the box that applies the most.)*

☐ very good	☐ good	☐ average	☐ poor
understands (almost) everything, knows a large number of words, makes only age-appropriate grammatical mistakes	often understands what is being said, knows many words, makes minor grammatical mistakes	understands the general idea of what is being said, knows some words, makes grammatical mistakes	understands and speaks only a few words, does not speak full sentences

B11) When speaking to your child, to what degree do you use German and your mother tongue? *(Please tick the box that applies the most.)*

☐	☐	☐	☐	☐
German almost never Mother tongue almost always	German seldom Mother tongue usually	German 50% Mother tongue 50%	German usually Mother tongue seldom	German almost always Mother tongue almost never

General Questions

C1) Compared to other Turkish-German bilingual children of the same age do you think your child's speech is intelligible? ☐ Yes ☐ No

C2) Do other people understand your child well when s/he speaks? ☐ Yes ☐ No

C3) Do people outside the family think your child has speech difficulties? ☐ Yes ☐ No

C4) Are you worried about your child's speech development? ☐ Yes ☐ No

Nursery Attendance

E1) When did your child enter the nursery? ______________________
month / year

E2) How regularly does your child attend the nursery?
☐ every day (Mondays to Fridays) ☐ four days a week ☐ three days a week
☐ two days a week ☐ one day per week ☐ less than one day per week

E3) How many hours does your child attend the nursery on a regular day? ☐

Thank you for completing this questionnaire!
Please return the questionnaire together with the Parents' Consent Form to the head nursery educator of your child's nursery.

Appendix E: Parents' questionnaire at t2

Speech Sound Development in Turkish-German Bilingual Children

Parents' Questionnaire

As you agree for your child to take part in the second assessment phase of this research project you are kindly asked to provide update information about the bilingual situation in your family, your child's use of Turkish and German and whether your child has received speech and language intervention in the meantime. This information will be treated as strictly confidential and will only be used to identify factors that may influence children's acquisition of speech sounds and their performance in the assessment.

As soon as we receive your completed questionnaire this first page containing your child's name will be destroyed and will be replaced by a research participant code on page 2. This questionnaire will be stored in a locked filing cabinet at the researcher's office together with the audio recordings. Only Katharina Salgert and her supervisors will be allowed access to this data which will be destroyed as soon as related publications are in press.

Your child's name: ____

Child's Code (to be filled in by the researcher): ______________________________

01) Who fills in the questionnaire? ☐ mother/female carer ☐ father/male carer

Use of Turkish

A1) Who speaks Turkish to your child?
☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

A2) Which dialectal variations of Turkish does your child hear on a regular basis (e.g., Istanbul-Turkish, Urfa, Dinler)?

____

A3) How regularly does your child **hear** Turkish?
☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

A4) When hearing Turkish, how many hours on average does your child **hear** Turkish per day? ☐

A5) My child speaks Turkish to:
☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

A6) How regularly does your child **speak** Turkish?
☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

A7) When speaking Turkish, how many hours on average does your child **speak** in Turkish per day? ☐

A8) Where does your child speak Turkish?
☐ at home ☐ in the nursery ☐ at friends' ☐ at relatives'
☐ at hobbies/leisure activities ☐ on holidays (in Turkey) ☐ in public

A9) How would you rate the proficiency of your child in Turkish? *(Please tick the box that applies the most.)*

☐ very good	☐ good	☐ average	☐ poor
understands (almost) everything, knows a large number of words, makes only age-appropriate grammatical mistakes	often understands what is being said, knows many words, makes minor grammatical mistakes	understands the general idea of what is being said, knows some words, makes grammatical mistakes	understands and speaks only a few words, does not speak full sentences

A10) How would you rate the proficiency of your child in Turkish? *(Please tick the box that applies the most.)*

☐ very good	☐ good	☐ average	☐ poor
understands (almost) everything, knows a large number of words, makes only age-appropriate grammatical mistakes	often understands what is being said, knows many words, makes minor grammatical mistakes	understands the general idea of what is being said, knows some words, makes grammatical mistakes	understands and speaks only a few words, does not speak full sentences

Acquisition and Use of German

B1) Who speaks German to your child?

☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

B2) When did your child start to hear German?

☐ from birth on when the child was: ☐ 1 year old ☐ 2 years old ☐ 3 years old
☐ 4 years old

B3) Does your child hear a different dialect al variation of German than the Hamburgian (Northern)?

☐ Yes If yes, please state which one: ______________________________
☐ No

B4) How regularly does your child **hear** German?

☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

B5) When hearing German, how many hours on average does your child **hear** German per day? ☐

B6) My child speaks German to:

☐ mother/female carer ☐ father/male carer ☐ siblings ☐ grandparents
☐ aunts/uncles ☐ neighbours ☐ friends ☐ nursery nurses

B7) How regularly does your child **speak** German?

☐ daily ☐ once or twice a week ☐ once a month ☐ seldom (e.g. on holiday)

B8) When speaking German, how many hours does your child **speak** in German on average per day? ☐

B9) Where does your child speak German?

☐ at home ☐ in the nursery ☐ at friends' ☐ at relatives'
☐ at hobbies/leisure activities ☐ on holidays (in Turkey) ☐ in public

General Questions

C1) Compared to other Turkish-German bilingual children of the same age do you think your child's speech is intelligible? ☐ Yes ☐ No

C2) Do other people understand your child well when s/he speaks? ☐ Yes ☐ No

C3) Do people outside the family think your child has speech difficulties? ☐ Yes ☐ No

C4) Are you worried about your child's speech development? ☐ Yes ☐ No

Nursery Attendance

D1) How regularly does your child attend the nursery?

☐ every day (Mondays to Fridays) ☐ four days a week ☐ three days a week
☐ two days a week ☐ one day per week ☐ less than one day per week

D2) How many hours does your child attend the nursery on a regular day? ☐

Speech and Language Therapy

E1) Did your child receive speech and language therapy in the time between the first assessment and today? ☐ Yes ☐ No

If yes…

a) Please indicate when the therapy started and when it ended.

Start: ____________________ End:______________________

b) Do you agree to the researcher contacting the SLT so that he can find out about the details of the intervention your child receives/d? This is particularly helpful to judge your child's speech abilities on the second assessment. ☐ Yes ☐ No

c) If yes, we would appreciate if you could also indicate the name of the clinic's owner.

Name:____________________________ Tel:_______________________

If no…

d) Is your child on a waiting list for a speech and language therapy place? ☐ Yes ☐ No

Thank you for completing this questionnaire!

Please return the questionnaire together with the Parents' Consent Form to the head nursery educator of your child's nursery.

Appendix F: SLT's responses to the interventions received between t1 and t2

Questions to SLTs	**Participant 1**	**Participant 2**
Therapy start	September 2014	October 2014
Therapy end	ongoing	ongoing
No. of sessions between t1 and t2	18	13
Focus of intervention	Phonology (CC-clusters) Expressive vocabulary Receptive syntax	Phonology Expressive syntax
Treatment approach	P.O.P.T.[1] (10 sessions) HOT[2] (8 sessions)	Articulation therapy[3] (10 sessions) HOT (3 sessions)
PA tasks included in therapy	Syllable segmentation Rime detection & production	Syllable segmentation
Language of the intervention	German	German

Note: **PA:** phonological awareness, **[1]**Psycholinguistic orientated phonology therapy (Fox, 2003), **[2]**Handlungsorientierter Therapieansatz [action orientated therapy approach] (Weigl & Reddemann-Tschaikner, 2001), **[3]**articulation therapy (van Riper, 1963)

Appendix G: Turkish Phonology Assessment – Items (incl. translation)

	IPA	Turkish	English
1	pɛjˈniɾ	Peynir	Cheese
2	ɟyˈnɛʃ	Güneş	Sun
3	bʌˈʊuɫ	Bavul	Suitcase
4	t͜ʃoˈd͜ʒuk	Çocuk	Child
5	dœɾt	Dört	Four
6	d͜ʒʌːˈmi	Cami	Mosque
7	kʌlp	Kalp	Heart
8	kʌhˈvɛ	Kahve	Coffee
9	lʌmˈbʌ	Lamba	Lamp
10	nɛzˈlɛ	Nezle	Sniffles/Cold
11	jʌːˈmuɾ	Yağmur	Rain
12	ɛʊ	Ev	House
13	diʃˈlɛɾ	Dişler	Teeth
14	sʌlɟʌnˈgoz	Salyangoz	Snail
15	mʌˈvi	Mavi	Blue
16	rɛnc	Renk	Colour
17	siˈjʌh	Siyah	Black
18	jɛˈʃil	Yeşil	Green
19	syt	Süt	Milk
20	fil	Fil	Elephant
21	gʌˈɾʌʒ	Garaj	Garage
22	dokˈtoɾ	Doktor	Doctor
23	itfʌiˈjɛ	Itfaiye	Fire Brigade
24	ʒœˈlɛ	Jöle	Jelly
25	ʃɛftʌˈli	Şeftali	Peach
26	ʃiʃˈmʌn	Şişman	Fat
27	zʌˈjɯf	Zayıf	Slim
28	joɾˈgʌn	Yorgan	Duvet
29	t͜ʃʌj	Çay	Tea
30	kʌɾˈpuz	Karpuz	Watermelon
31	zyɾʌˈfʌ	Zürafa	Giraffe
32	piˈɾint͜ʃ	Pirinç	Rice (raw)
33	pʌɾc	Park	Play Ground
34	sʌɫɯnˈd͜ʒʌk	Salıncak	Swing
35	sʌndʌɫˈjɛ	Sandalye	Chair
36	vʌˈgon	Vagon	Coach
37	ɛldiˈvɛn	Eldiven	Glove
38	yt͜ʃˈɟɛn	Üçgen	Triangle
39	hɛdiˈjɛ	Hediye	Present
40	d͜ʒɛˈcɛt	Ceket	Jacket
41	tʌˈʊuk	Tavuk	Chicken
42	vʌk (vʌk)	vak vak	Quack quack
43	t͜ʃʌˈtʌɫ	Çatal	Fork
44	kuʃ	Kuş	Bird
45	iˈci	Iki	Two
46	ʌːt͜ʃ	Ağaç	Tree
47	mʌˈsʌ	Masa	Table
48	kuɾˈbʌː	Kurbağa	Frog
49	pʌsˈtʌ	Pasta	Cake
50	ɟɛˈmi	Gemi	Ship
51	ɾyzˈɟʌɾ	Rüzgar	Wind
52	ʌnʌhˈtʌɾ	Anahtar	Key
53	bʌˈɫɯk	Balık	Fish
54	hʌˈʊut͜ʃ	Havuç	Carrot
55	cœˈpɛc	Köpek	Dog
56	hʌʊ (hʌʊ)	hav hav	Woof woof
57	zɛjˈtin	Zeytin	Olive
58	gʌzɛˈtɛ	Gazete	Newspaper
59	ɛcˈmɛc	Ekmek	Bread
60	ʃʌpˈkʌ	Şapka	Hat
61	ɟœzˈlyc	Gözlük	Glasses
62	mʌˈkʌs	Makas	Scissors
63	yˈzym	Üzüm	Grape
64	t͜ʃiftˈlic	Çiftlik	Farm
65	bʌˈd͜ʒʌk	Bacak	Leg
66	ʒimnʌsˈtic	Jimnastik	Gymnastics
67	fʌˈɾɛ	Fare	Mouse
68	ɾɛˈt͜ʃɛl	Reçel	Jam
69	citˈʌp	Kitap	Book
70	ɛʒdɛɾˈhʌ	Ejderha	Dragon

Appendix H: Turkish Phonology Assessment – Record form

Turkish Phonology Assessment

for Turkish-German Bilingual Children (Albrecht, in prep.)

Child: ____________________ Date: ____________________ Folder/Track No.: [/]

(Code)

T: ___

Instructions:

Bak bi, sana resimler getirdim ve sen kesin hepsini taniyorsundur. Gel, birde beraber bakalım.
Ne görüyorsun? / Bu ne? / İsmi ne?
Bunun Türkçesi ne?

Example for prompts:

(E) *Bunu kağit kesmek isteyince kullaniyorsun. Bu bir....?*

(C) *Bu <u>süt mü</u> veya su mu?* (target word first!)

(I) *Bu kuş. Sen bir "kuş" dermisin?*

Documentation:

- Please tick which prompts you used (**E**: Explanation of the item's function, **C**: Choice, **I**: Imitation)
- Please make a tick when the pronunciation of the target word was fully correct.
- In case the child's pronunciation differs from the target pronunciation, write down the child's realisation in IPA.

No.	Item	IPA	Prompts	Child's Realisation	Patterns
1	Peynir	pɛjˈniɾ	(E) (C) (I)		
2	Güneş	ɟyˈnɛʃ	(E) (C) (I)		
3	Bavul	bʌˈʊuɫ	(E) (C) (I)		
4	Çocuk	t͜ʃoˈd͜ʒuk	(E) (C) (I)		
Kaç yıldız görüyorsun?					
5	Dört	dœɾt	(E) (C) (I)		
6	Cami	d͜ʒʌːˈmi	(E) (C) (I)		
7	Kalp	kʌlp	(E) (C) (I)		
8	Kahve	kʌhˈvɛ	(E) (C) (I)		
9	Lamba	lʌmˈbʌ	(E) (C) (I)		
10	Nezle	nɛzˈlɛ	(E) (C) (I)		
11	Yağmur	jʌːˈmuɾ	(E) (C) (I)		

12	Ev	ɛʊ	(E) (C) (I)		
13	Dişler	diʃˈlɛr	(E) (C) (I)		
14	Salyangoz	sʌɫjʌnˈgoz	(E) (C) (I)		
Hangi renk?					
15	Mavi	mʌˈvi	(E) (C) (I)		
Mavi nedir …?					
16	Renk	rɛnc	(E) (C) (I)		
Hangi renk?					
17	Siyah	siˈjʌh	(E) (C) (I)		
Hangi renk?					
18	Yeşil	jɛˈʃil	(E) (C) (I)		
19	Süt	syt	(E) (C) (I)		
20	Fil	fil	(E) (C) (I)		
21	Garaj	gʌˈɾʌʒ	(E) (C) (I)		
22	Doktor	dokˈtor	(E) (C) (I)		
23	Itfaiye	itfʌiˈjɛ	(E) (C) (I)		
24	Jöle	ʒœˈlɛ	(E) (C) (I)		
25	Şeftali	ʃɛftʌˈli	(E) (C) (I)		
Bu adam nasıl?					
26	Şişman	ʃiʃˈmʌn	(E) (C) (I)		
Bu bayan nasıl?					
27	Zayıf	zʌˈjɯf	(E) (C) (I)		
28	Yorgan	jorˈgʌn	(E) (C) (I)		
29	Çay	t͜ʃʌj	(E) (C) (I)		
30	Karpuz	kʌɾˈpuz	(E) (C) (I)		
31	Zürafa	zyɾʌˈfʌ	(E) (C) (I)		
32	Pirinç	piˈɾint͜ʃ	(E) (C) (I)		
33	Park	pʌrc	(E) (C) (I)		

34	Salıncak	sʌɫɯnˈd͜ʒʌk	Ⓔ Ⓒ Ⓘ		
35	Sandalye	sʌndʌɫ̟ˈjɛ	Ⓔ Ⓒ Ⓘ		
36	Vagon	vʌˈgon	Ⓔ Ⓒ Ⓘ		
37	Eldiven	ɛldiˈvɛn	Ⓔ Ⓒ Ⓘ		
38	Üçgen	yt͜ʃˈɟɛn	Ⓔ Ⓒ Ⓘ		
39	Hediye	hɛdiˈjɛ	Ⓔ Ⓒ Ⓘ		
40	Ceket	d͜ʒɛˈcɛt	Ⓔ Ⓒ Ⓘ		
41	Tavuk	tʌˈʋuk	Ⓔ Ⓒ Ⓘ		
Ördekler nasıl ses çıkarır?					
42	vak vak	vʌk (vʌk)	Ⓔ Ⓒ Ⓘ		
43	Çatal	t͜ʃʌˈtʌɫ̟	Ⓔ Ⓒ Ⓘ		
44	Kuş	kuʃ	Ⓔ Ⓒ Ⓘ		
Kaç yıldız görüyorsun?					
45	Iki	iˈci	Ⓔ Ⓒ Ⓘ		
46	Ağaç	ʌ:t͜ʃ	Ⓔ Ⓒ Ⓘ		
47	Masa	mʌˈsʌ	Ⓔ Ⓒ Ⓘ		
48	Kurbağa	kuɾˈbʌ:	Ⓔ Ⓒ Ⓘ		
49	Pasta	ˈpʌstʌ	Ⓔ Ⓒ Ⓘ		
50	Gemi	ɟɛˈmi	Ⓔ Ⓒ Ⓘ		
51	Rüzgar	ɾyzˈɟʌɾ	Ⓔ Ⓒ Ⓘ		
52	Anahtar	ʌnʌhˈtʌɾ	Ⓔ Ⓒ Ⓘ		
53	Balık	bʌˈɫɯk	Ⓔ Ⓒ Ⓘ		
54	Havuç	hʌˈʋut͜ʃ	Ⓔ Ⓒ Ⓘ		
55	Köpek	cœˈpɛc	Ⓔ Ⓒ Ⓘ		
Köpekler nasıl ses çıkarır?					
56	hav hav	hʌʋ (hʌʋ)	Ⓔ Ⓒ Ⓘ		
57	Zeytin	zɛjˈtin	Ⓔ Ⓒ Ⓘ		

58	Gazete	gʌzɛˈtɛ	Ⓔ Ⓒ Ⓘ		
59	Ekmek	ɛcˈmɛc	Ⓔ Ⓒ Ⓘ		
60	Şapka	ʃʌpˈkʌ	Ⓔ Ⓒ Ⓘ		
61	Gözlük	ɟœzˈlyc	Ⓔ Ⓒ Ⓘ		
62	Makas	mʌˈkʌs	Ⓔ Ⓒ Ⓘ		
63	Üzüm	yˈzym	Ⓔ Ⓒ Ⓘ		
64	Çiftlik	t͜ʃiftˈlic	Ⓔ Ⓒ Ⓘ		
65	Bacak	bʌˈd͜ʒʌk	Ⓔ Ⓒ Ⓘ		
66	Jimnastik	ʒimnʌsˈtic	Ⓔ Ⓒ Ⓘ		
67	Fare	fʌˈrɛ	Ⓔ Ⓒ Ⓘ		
68	Reçel	rɛˈt͜ʃɛl	Ⓔ Ⓒ Ⓘ		
69	Kitap	ciˈtʌp	Ⓔ Ⓒ Ⓘ		
70	Ejderha	ɛʒdɛrˈhʌ	Ⓔ Ⓒ Ⓘ		

Appendix I: Turkish Phonology Assessment – Items ordered by consonants and CCs assessed (incl. common allophones)

Phoneme	Word-initial	Syllable-initial	Syllable-final	Word-final	F	needs to occur to 2/3 of F (rounded numbers)
m	mʌˈkʌs mʌˈsʌ mʌˈvi	ɟeˈmi ɛcˈmɛc ʃiʃˈmʌn d͜ʒʌːˈmi jʌːˈmuɾ	lʌmˈbʌ ʒimnʌsˈtic	yˈzym	**11**	**7**
n	nɛzˈlɛ	ʌnʌhˈtʌɾ pɛjˈniɾ ɟyˈnɛʃ ʒimnʌsˈtic	sʌndʌɫˈjɛ sʌɫɯnˈd͜ʒʌk sʌɫɟʌnˈgoz	ʃiʃˈmʌn joɾˈgʌn ɛldiˈvɛn yt͜ʃˈɟɛn zɛjˈtin vʌˈgon	**15**	**10**
p	pʌsˈtʌ pɛjˈniɾ piˈɾint͜ʃ pʌɾc	kʌɾˈpuz cœˈpɛc	ʃʌpˈkʌ	citˈʌp	**8**	**5**
b	bʌˈɫɯk bʌˈd͜ʒʌk bʌˈʊuɫ	kuɾˈbʌː lʌmˈbʌ	---	---	**5**	**3**
t	tʌˈʊuk	gʌzɛˈtɛ ʌnʌhˈtʌɾ t͜ʃʌˈtʌɫ pʌsˈtʌ citˈʌp ʒimnʌsˈtic ʃɛftʌˈli dokˈtoɾ zɛjˈtin	itfʌiˈjɛ	syt d͜ʒɛˈcɛt	**13**	**9**
d	dokˈtoɾ dœɾt diʃˈlɛɾ	hɛdiˈjɛ sʌndʌɫˈjɛ ɛldiˈvɛn ɛʒdɛɾˈhʌ	---	---	**7**	**5**
c	cœˈpɛc citˈʌp	iˈci d͜ʒɛˈcɛt	ɛcˈmɛc	ɛcˈmɛc cœˈpɛc ɟœzˈlyc ʒimnʌsˈtic t͜ʃiftˈlic	**10**	**7**
k	kuʃ kuɾˈbʌː kʌɾˈpuz kʌhˈvɛ kʌlp	mʌˈkʌs ʃʌpˈkʌ	dokˈtoɾ	tʌˈʊuk bʌˈɫɯk sʌɫɯnˈd͜ʒʌk bʌˈd͜ʒʌk vʌk (vʌk) t͜ʃoˈd͜ʒuk	**14**	**9**
ɟ	ɟɛˈmi ɟœzˈlyc ɟyˈnɛʃ	yt͜ʃˈɟɛn ɾyzˈɟʌɾ	---	---	**5**	**3**
g	gʌzɛˈtɛ	sʌɫɟʌnˈgoz	---	---	**5**	**3**

Phoneme	Word-initial	Syllable-initial	Syllable-final	Word-final	F	needs to occur to 2/3 of F (rounded numbers)
	gʌˈɾʌʒ	jorˈgʌn vʌˈgon				
f	fil fʌˈɾɛ	zyɾʌˈfʌ itfʌiˈjɛ	ʃɛftʌˈli	zʌˈjɯf	**6**	**4**
v	vʌk (vʌk) vʌˈgon	mʌˈvi kʌhˈvɛ ɛldiˈvɛn		---	**5**	**3**
ʊ	---	tʌˈʊuk hʌˈʊut͜ʃ bʌˈʊuɫ	---	ɛʊ hʌʊ (hʌʊ)	**5**	**3**
s	syt sʌɫjʌnˈgoz siˈjʌh sʌndʌɫˈjɛ sʌɫɯnˈd͜ʒʌk	mʌˈsʌ	pʌsˈtʌ ʒimnʌsˈtic	mʌˈkʌs	**9**	**6**
z	zyɾʌˈfʌ zɛjˈtin zʌˈjɯf	yˈzym gʌzɛˈtɛ	ɟœzˈlyc ɾyzˈɟʌɾ nɛzˈlɛ	kʌɾˈpuz sʌɫjʌnˈgoz	**10**	**7**
ʃ	ʃʌpˈkʌ ʃiʃˈmʌn ʃɛftʌˈli	jɛˈʃil	ʃiʃˈmʌn diʃˈlɛɾ	ɟyˈnɛʃ kuʃ	**8**	**5**
ʒ	ʒimnʌsˈtic ʒœˈlɛ		ɛʒdɛɾˈhʌ	gʌˈɾʌʒ	**4**	**3**
(ɣ)	---		ʌːt͜ʃ kuɾˈbʌː jʌːˈmuɾ		**3**	**2**
h	hɛdiˈjɛ hʌˈʊut͜ʃ hʌʊ (hʌʊ)	ɛʒdɛɾˈhʌ	ʌnʌhˈtʌɾ kʌhˈvɛ	siˈjʌh	**7**	**5**
ɾ	ɾɛˈt͜ʃɛl ɾyzˈɟʌɾ ɾɛnc	zyɾʌˈfʌ gʌˈɾʌʒ fʌˈɾɛ piˈɾint͜ʃ	kuɾˈbʌː kʌɾˈpuz jorˈgʌn ɛʒdɛɾˈhʌ	ʌnʌhˈtʌɾ pɛjˈniɾ jʌːˈmuɾ dokˈtoɾ ɾyzˈɟʌɾ diʃˈlɛɾ	**17**	**11**
l	lʌmˈbʌ	ɟœzˈlyc ʒœˈlɛ ʃɛftʌˈli t͜ʃiftˈlic diʃˈlɛɾ nɛzˈlɛ	ɛldiˈvɛn	fil jɛˈʃil ɾɛˈt͜ʃɛl	**11**	**7**
ɫ	---	sʌɫɯnˈd͜ʒʌk bʌˈɫɯk	sʌndʌɫˈjɛ sʌɫjʌnˈgoz	t͜ʃʌˈtʌɫ bʌˈʊuɫ	**6**	**4**
j	jɛˈʃil jʌːˈmuɾ jorˈgʌn	hɛdiˈjɛ siˈjʌh sʌndʌɫˈjɛ sʌɫjʌnˈgoz itfʌiˈjɛ zʌˈjɯf	pɛjˈniɾ zɛjˈtin	t͜ʃʌj	**12**	**8**

Phoneme	Word-initial	Syllable-initial	Syllable-final	Word-final	F	needs to occur to 2/3 of F (rounded numbers)
tʃ	tʃʌj tʃʌˈtʌɫ tʃiftˈlic tʃoˈdʒuk	rɛˈtʃɛl	ytʃˈɟɛn	ʌ: tʃ hʌˈʊutʃ	**8**	**5**
dʒ	dʒʌ:ˈmi dʒɛˈcɛt	sʌɫɯnˈdʒʌk bʌˈdʒʌk tʃoˈdʒuk	---	---	**5**	**3**

Note: **---**: Phoneme does not occur in this position, Phoneme does not occur in child appropriate words in this position, **(ɣ)**: There is an on-going debate about how the Turkish letter <ğ> is pronounced and transcribed. Kopkallı-Yavuz (2010) and Topbaş (2007) argue that it is not pronounced at all in Modern Turkish and only functions to lengthen the preceding vowel (i.e., [ʌ:tʃ], [kuɾˈbʌ:], [jʌ:ˈmuɾ]). However, it does occur in some non-standard dialects as a voiced velar fricative (i.e., [ʌɣˈʌtʃ], [kuɾbʌɣˈʌ], [jʌɣˈmuɾ]; Topbaş & Yavaş, 2006). The items I have included with this "sound" are primarily included to assess other sounds in the respective words and only secondarily to investigate the presence of <ğ>, **F**: Frequency

Cluster	Acquisition of CCs (Topbaş 2007)	Syllable final	Word final
ntʃ	Early		piˈri**ntʃ**
nc			rɛ**nc**
ft	Middle	tʃi**ft**ˈlic	
st			
lp			kʌ**lp**
lk			
ɾt	Late		dœ**ɾt**
ɾc			pʌ**ɾc**
ɾf			

Note: Phoneme does not occur in child appropriate words in this position.

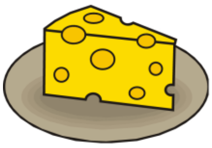
Appendix J: Turkish Phonology Assessment – Items ordered by vowel phonemes assessed (according to Kopkallı-Yavuz, 2010)

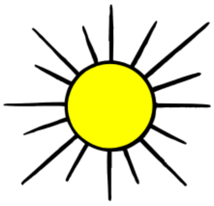
Vowel	**Items**					**Frequency**
ʌ	bʌˈɫɯk	kurˈbʌ:	bʌˈd͜ʒʌk	bʌˈʋuɫ	lʌmˈbʌ	**58 (56)**
	pʌsˈtʌ	ʌ:t͜ʃ	sʌndʌɫˈjɛ	ɛʒdɛrˈhʌ	fʌˈrɛ	
	itfʌiˈjɛ	zyrʌˈfʌ	ʃɛftʌˈli	zʌˈjɯf	gʌzɛˈtɛ	
	gʌˈrʌʒ	jorˈgʌn	sʌɫjʌnˈgoz	vʌˈgon	ryzˈɟʌr	
	hʌˈʋut͜ʃ	hʌʋ (hʌʋ)	mʌˈsʌ	siˈjʌh	ʌnʌhˈtʌr	
	kʌhˈvɛ	jʌ:ˈmur	t͜ʃʌj	kʌrˈpuz	ʃʌpˈkʌ	
	kʌlp	mʌˈkʌs	vʌk (vʌk)	tʌˈʋuk	sʌɫɯnˈd͜ʒʌk	
	ciˈtʌp	ʒimnʌsˈtic	d͜ʒʌ:ˈmi	t͜ʃʌˈtʌɫ	ʃiʃˈmʌn	
	mʌˈvi	pʌrc				
ɛ	fʌˈrɛ	diʃˈlɛr	ɛldiˈvɛn	hɛdiˈjɛ	rɛnc	**32**
	sʌndʌɫˈjɛ	ɛʒdɛrˈhʌ	itfʌiˈjɛ	nɛzˈlɛ	ʃɛftʌˈli	
	ɛʋ	gʌzɛˈtɛ	ɟɛˈmi	ɟyˈnɛʃ	yt͜ʃˈɟɛn	
	kʌhˈvɛ	jɛˈʃil	zɛjˈtin	pɛjˈnir	cœˈpɛc	
	d͜ʒɛˈcɛt	ɛcˈmɛc	ʒœˈlɛ	rɛˈt͜ʃɛl		
i	diʃˈlɛr	ɛldiˈvɛn	hɛdiˈjɛ	mʌˈvi	fil	**24**
	itfʌiˈjɛ	ʃɛftʌˈli	ɟɛˈmi	siˈjʌh	jɛˈʃil	
	zɛjˈtin	pɛjˈnir	ciˈtʌp	iˈci	ʒimnʌsˈtic	
	t͜ʃiftˈlic	piˈrint͜ʃ	ʃiʃˈmʌn	d͜ʒʌ:ˈmi		
o	dokˈtor	vʌˈgon	jorˈgʌn	sʌɫjʌnˈgoz	t͜ʃoˈd͜ʒuk	**6**
u	bʌˈʋuɫ	kurˈbʌ:	hʌˈʋut͜ʃ	jʌ:ˈmur	kuʃ	**8**
	kʌrˈpuz	tʌˈʋuk	t͜ʃoˈd͜ʒuk			
y	zyrʌˈfʌ	ɟyˈnɛʃ	ɟœzˈlyc	yt͜ʃˈɟɛn	ryzˈɟʌr	**8**
	yˈzym	syt				
œ	dœrt	ʒœˈlɛ	ɟœzˈlyc	cœˈpɛc		**4**
ɯ	bʌˈɫɯk	sʌɫɯnˈd͜ʒʌk	zʌˈjɯf			**3**

Appendix K: Turkish Phonology Assessment – Items ordered by number of syllables assessed

monosyllabic	disyllabic	trisyllabic
ʌ:t͜ʃ	bʌˈɫɯk	sʌndʌɫˈjɛ
hʌʊ (hʌʊ)	kuɾˈbʌ:	ɛʒdɛɾˈhʌ
t͜ʃʌj	bʌˈd͜ʒʌk	itfʌiˈjɛ
kʌlp	bʌˈʊuɫ	zyɾʌˈfʌ
vʌk (vʌk)	lʌmˈbʌ	ʃɛftʌˈli
ɾɛnc	fʌˈɾɛ	gʌzɛˈtɛ
ɛʊ	gʌˈɾʌʒ	sʌɫjʌnˈgoz
fil	joɾˈgʌn	ʌnʌhˈtʌɾ
kuʃ	vʌˈgon	sʌɫɯnˈd͜ʒʌk
syt	ɾyzˈɟʌɾ	ʒimnʌsˈtic
pʌɾc	hʌˈʊut͜ʃ	ɛldiˈvɛn
dœɾt	siˈjʌh	hɛdiˈjɛ
	kʌhˈvɛ	
	jʌ:ˈmuɾ	
	kʌɾˈpuz	
	ʃʌpˈkʌ	
	mʌˈkʌs	
	tʌˈʊuk	
	ciˈtʌp	
	t͜ʃʌˈtʌɫ	
	ʃiʃˈmʌn	
	d͜ʒʌˈmi	
	mʌˈsʌ	
	pʌsˈtʌ	
	mʌˈvi	
	diʃˈlɛɾ	
	ɟɛˈmi	
	ɟyˈnɛʃ	
	yt͜ʃˈɟɛn	
	jɛˈʃil	
	zɛjˈtin	
	pɛjˈniɾ	
	cœˈpɛc	
	d͜ʒɛˈcɛt	
	ɛcˈmɛc	
	ʒœˈlɛ	
	ɾɛˈt͜ʃɛl	
	nɛzˈlɛ	
	iˈci	
	piˈɾint͜ʃ	
	dokˈtoɾ	
	t͜ʃoˈd͜ʒuk	
	yˈzym	
	t͜ʃiftˈlic	
	ɟœzˈlyc	
	zʌˈjɯf	
12	**46**	**12**

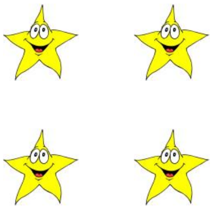
Appendix L: Turkish Phonology Assessment – Pictures

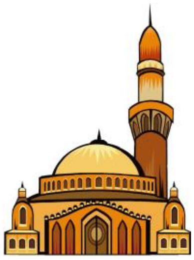

1


2


3


4


5


6


7


8


9


10


11

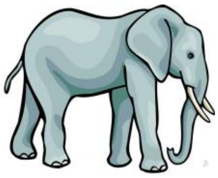

12


13


14


15, 16, 17, 18

19


20

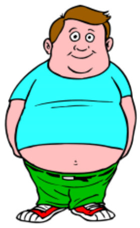

21


22


23

24

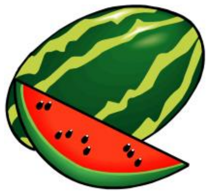

25


26


27


28

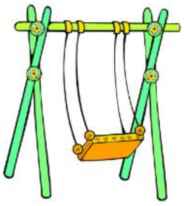

29


30


31


32

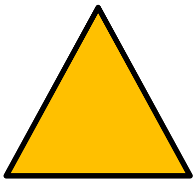

33


34


35

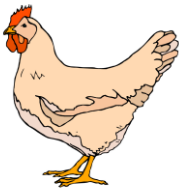

36

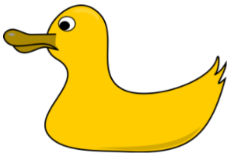

37


38


39

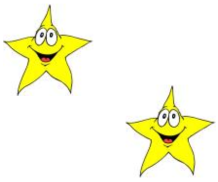

40


41

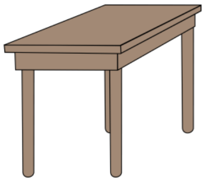

42

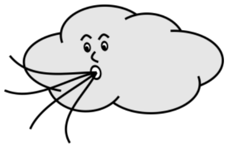
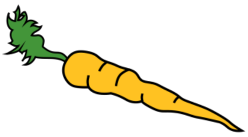
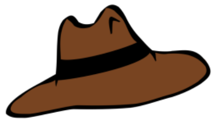
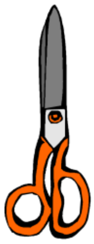
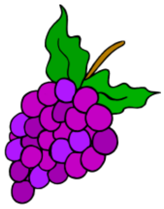
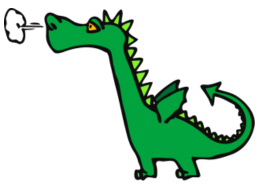

43


44


45

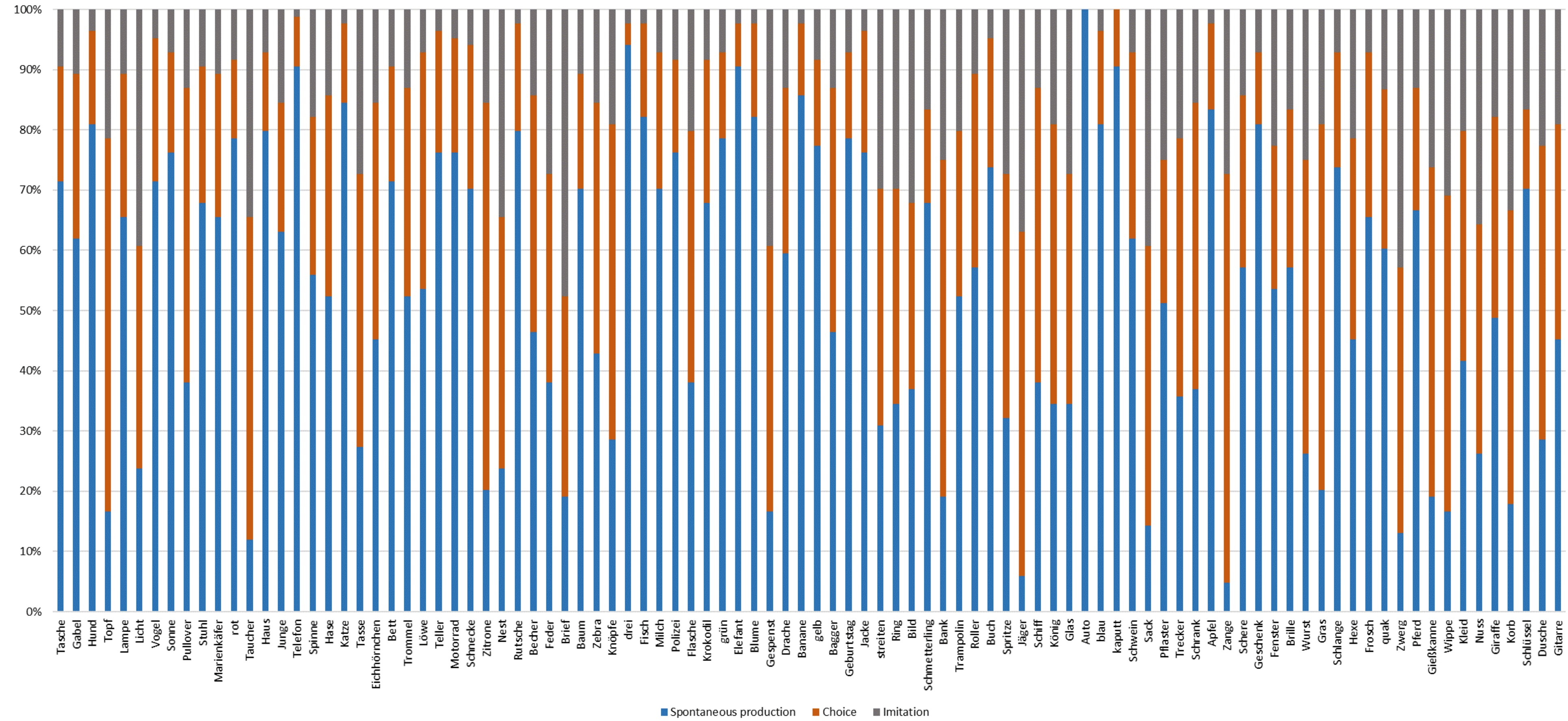

46


47

 48	 49	 50	 51
 52	 53	 55	 5556
 57	 58	 59	 60
 61	 62	 63	 64
 65	 66	 67	 68
 69	 70		

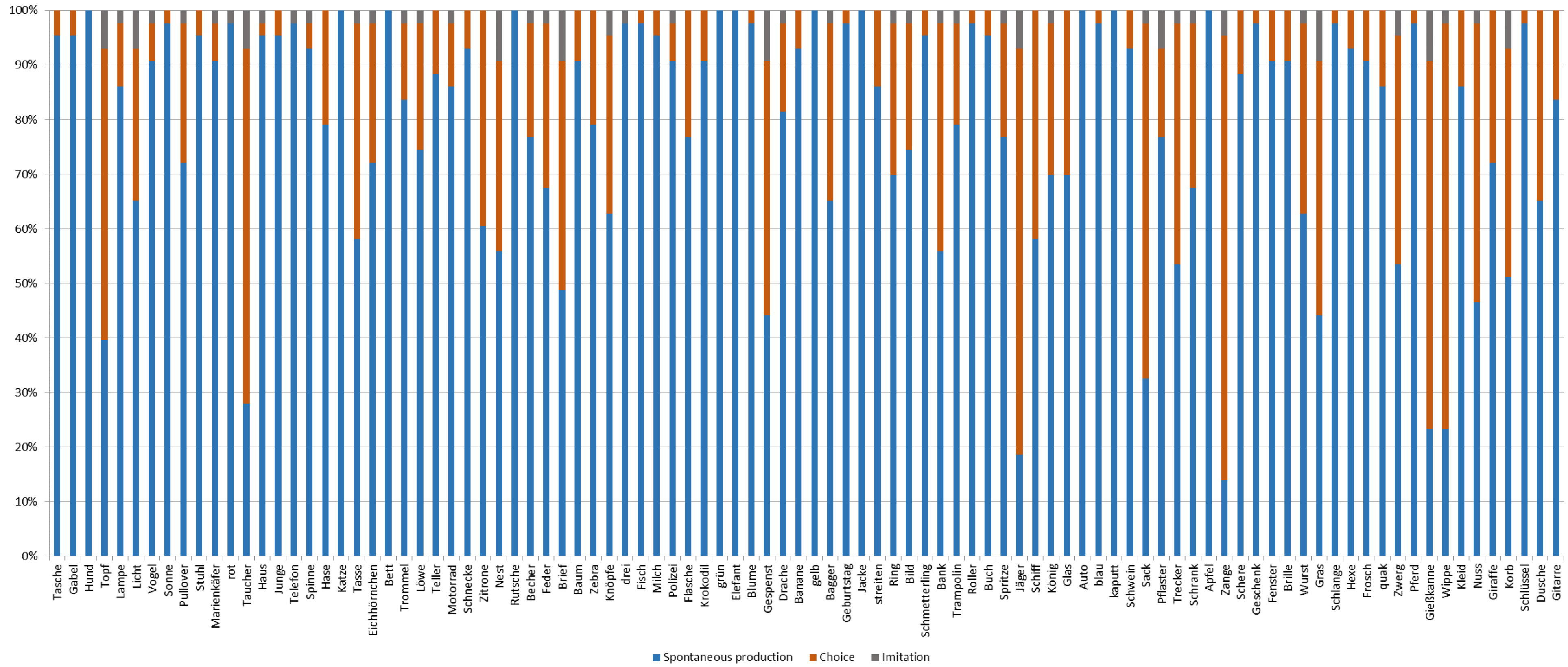
Reference	Pictures
http://www.clipart.com accessed on: 25/01/2011	1, 2, 4, 5, 6, 7, 9, 11, 12, 14, 19, 20, 26, 30, 31, 34, 35, 39, 40, 41, 43, 44, 45, 46, 48, 49, 50, 52, 53, 58, 59, 61, 62, 63, 68, 69
http://www.openclipart.org accessed on: 28/10/2012	3, 8, 10, 15, 16, 17, 18, 21, 22, 23, 25, 28, 29, 36, 37, 38, 42, 47, 51, 54, 55, 56, 60, 65, 67, 70
http://office.microsoft.comen-usimages?CTT=97 accessed on: 28/10/2012	33, 66
http://www.clker.com accessed on: 23/11/2012	13, 27, 32
Türkisch Artikulationstest - TAT (Naş, 2010)	57
http://www.kenton.com.trblogbitkisel-joleler.html accessed on: 25/03/2013	24
http://www.animaatjes.demalvorlagenbauernhof accessed on: 26/03/2013	64

Appendix M: Percentage of children who needed a certain prompt to name the items in the PLAKSS-II at t1



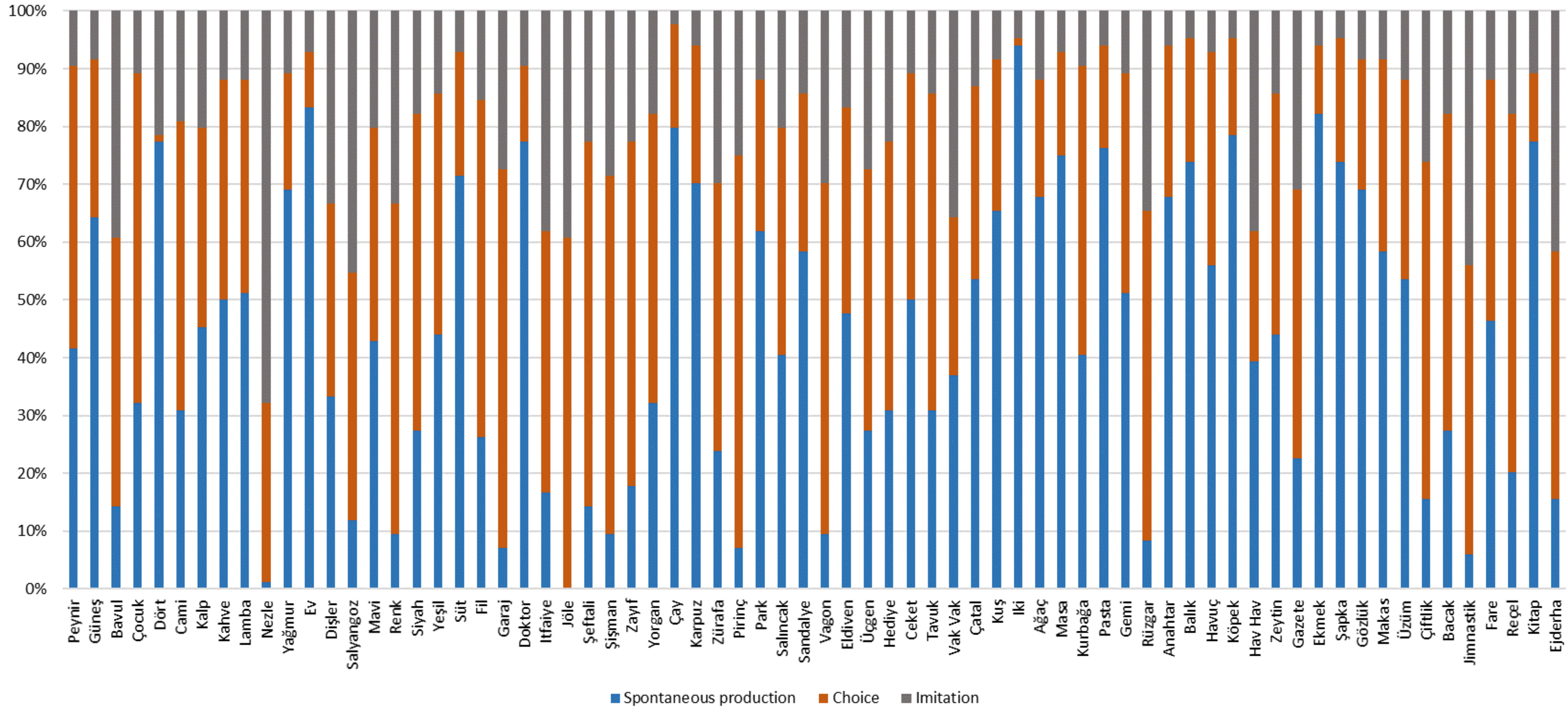
Note: The items are displayed in the order of their occurrence in the test.

Appendix N: Percentage of children who needed a certain prompt to name the items in the PLAKSS-II at t2



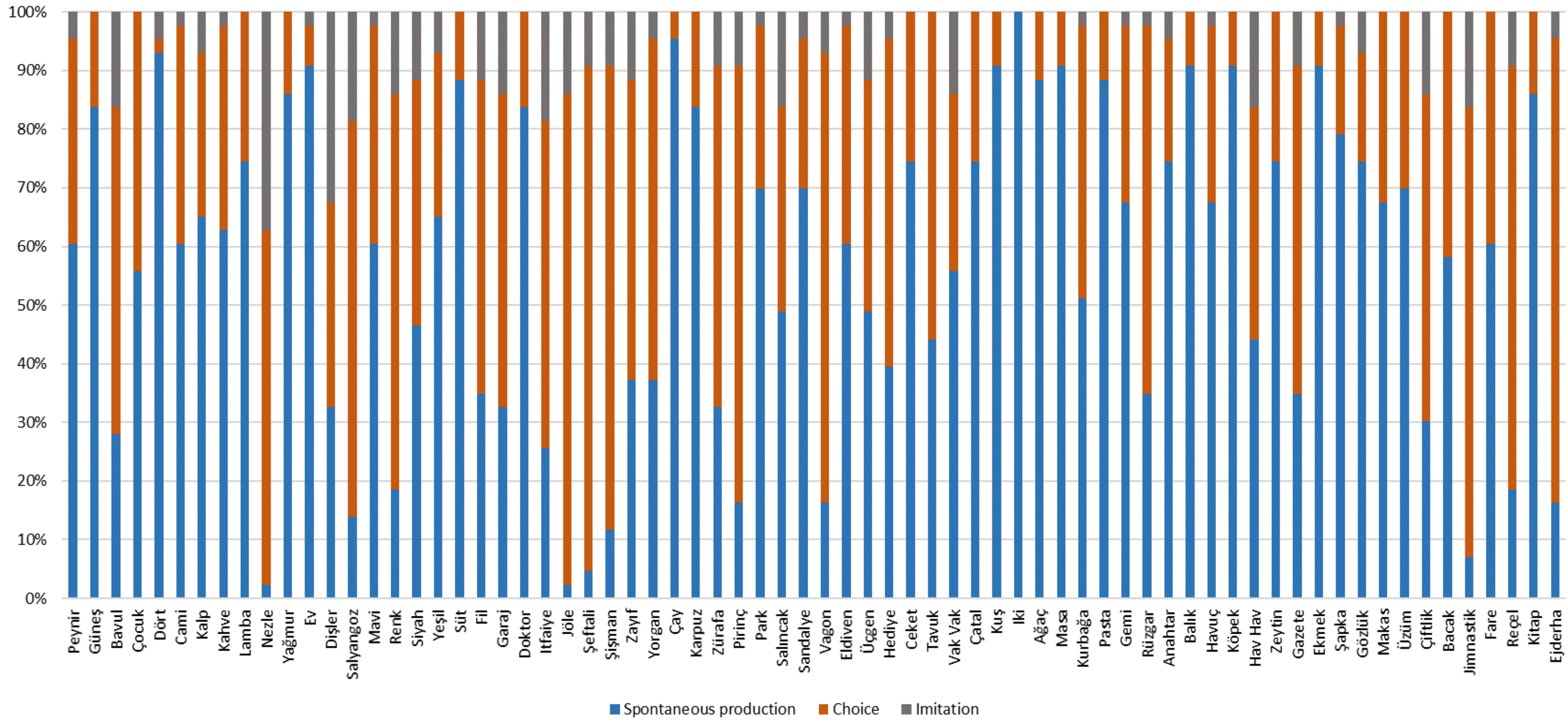
Note: The items are displayed in the order of their occurrence in the test.

Appendix O: Percentage of children who needed a certain prompt to name the items in the TPA at t1



Note.:The items are displayed in the order of their occurrence in the test.

Appendix P: Percentage of children who needed a certain prompt to name the items in the TPA at t2



Note: The items are displayed in the order of their occurrence in the test.

Appendix Q: Instructions for the administration of the non-word repetition and consistency task

Instruction:

I have brought a little teddy-game with me on my laptop.

Look, here is the teddy and there is a pot of honey. The little teddy would like to eat the honey because teddies love honey. But we need to help the teddy so that he can get the honey.

The teddy is now going to say a nonsense word. A nonsense word is a word that does not exist and which sounds funny. It is like a secret language. So, when the teddy says the nonsense word, you listen very carefully and then repeat the word. After that you say the word again three times. I will count the times with my fingers so that you can always see for how many times you still need to say the word.

Let us practice this now.

Play the first practice item and then ask the child: Could you repeat what the teddy has just said?

Please note all prompts you have given during the practice items on the record form. Also, take a note of repetitions and self-corrections made by the child.

Continue with the next practice item and then add: Well done! That is what we will do with all the other boxes. The teddy is going to say a nonsense word you repeat it and then say it again three times. As soon as we have finished all the boxes we will find out whether the teddy will get the honey, alright?

One repetition per item is allowed only if:

- *the child demands it.*
- *the child was not paying attention distracted.*
- *the child started talking as soon as the audio was played, thus resulting in an overlap.*

Appendix R: Record sheet NWR & consistency t2

Participant: _____________

Date: ___________ Age: ________

Non-word Repetition

No.	Item	Transcription	Child's realisation	Phonol. Patterns
P1	**Pilu**	[ˈpilu]		
Feedback/Notes				
P2	**Fulpi**	[ˈfulpi]		
Feedback/Notes				
1	Faku	[ˈfaːku]		
2	Fappusch	[ˈfapuʃ]		
3	Pilfu	[ˈpilfu]		
4	Kifapu	[kiˈfaːpu]		
5	Kupalfi	[kuˈpalfi]		
6	Pufaki	[puˈfaːki]		
7	Pufischkufi	[pufiʃˈkuːfi]		
8	Kalfipaku	[kalfiˈpaːku]		
9	Schufikapi	[ʃufiˈkaːpi]		
10	Kischupifaku	[kiʃuˈpiːfaku]		
11	Fapukaschifal	[fapuˈkaːʃifal]		
12	Puschifakilpa	[puʃiˈfaːkilpa]		
		Score correct (strict)		
		Score correct (lenient)		

Consistency

No.	1st repetition	2nd repetition	3rd repetition	S
P1				
Feedback/Notes				
P2				
Feedback/Notes				
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				

S: scoring

Appendix S: Possible phonetic and dialectal/accentual realisations of the PLAKSS-II-items not scored as incorrect

Items PLAKSS- II	Standard German	Accepted phonetic/dialectal/accentual variations
Tasche	ˈtaʃə	
Gabel	ˈgabəl	/bəl/ → [bl]
Hund	hʊnt	
Topf	tɔp͜f	
Lampe	ˈlampə	
Licht	lɪçt	/ç/ → [ʃ]
Vogel	ˈfo:gəl	/gəl/ → [gl]
Sonne	ˈzɔnə	/z/ → [ð]
Pullover	pʊˈlovɐ	
Stuhl	ʃtu:l	/ʃt/ → [ʃ + V+ t]
Marienkäfer	maˈʁi:nkɛ:fɐ	/ʁ/ → [ɾ, r, j, l]
rot	ʁot	/ʁ/ → [ɾ, r, j, l]
Taucher	ˈtaʊxɐ	
Haus	haʊs	
Junge	ˈjʊŋə	/ŋ/ → [ng, ŋg]
Telefon	ˈte:ləfo:n	
Spinne	ˈʃpɪnə	/ʃp/ → [ʃ + V + p]
Hase	ˈha:zə	/z/ → [ð]
Katze	ˈkat͜sə	/t͜s/ → [t͜θ]
Tasse	ˈtasə	/s/ → [θ]
Eichhörnchen	ˈaɪçhø:ɐnçən	/ç/ → [ʃ]
Bett	bɛt	
Trommel	ˈtʁɔməl	/ʁ/ → [ɾ, r, j, l], /tʁ/ → [t + V + ʁ]
Löwe	ˈlø:və	
Teller	ˈtɛlɐ	
Motorrad	moˈto:ʁat	/ʁ/ → [ɾ, r, j, l]
Schnecke	ˈʃnɛkə	/ʃn/ → [ʃ + V + n]
Zitrone	t͜siˈtʁo:nə	/t͜s/ → [s], /t͜s/ → [t͜θ], /ʁ/ → [ɾ, r, j, l], /tʁ/ → [t + V + ʁ]
Nest	nɛst	/s/ → [θ]
Rutsche	ˈʁʊtʃə	/ʁ/ → [ɾ, r, j, l]
Becher	ˈbɛçɐ	/ç/ → [ʃ]
Feder	ˈfe:dɐ	
Brief	bʁi:f	/ʁ/ → [ɾ, r, j, l], /bʁ/ → [b + V + ʁ]
Baum	baʊm	
Zebra	ˈt͜se:bʁa	/t͜s/ → [s], /t͜s/ → [t͜θ], /ʁ/ → [ɾ, r, j, l], /bʁ/ → [b + V+ ʁ]
Knöpfe	ˈknœp͜fə	/kn/ → [k + V + n]
Drei	dʁaɪ	/ʁ/ → [ɾ, r, j, l], /dʁ/ → [d + V+ ʁ]
Fisch	fɪʃ	

Milch	mɪlç	/ç/ → [ʃ]
Polizei	poliˈt͜saɪ	/t͜s/ → [t͜θ]
Flasche	ˈflaʃə	/fl/ → [f + V + l]
Krokodil	kʁokoˈdiːl	/ʁ/ → [ɾ, r, j, l], /kʁ/ → [k + V + ʁ]
grün	gʁyːn	/ʁ/ → [ɾ, r, j, l], /gʁ/ → [g + V + ʁ]
Elefant	eləˈfant	
Blume	ˈbluːmə	/bl/ → [b + V + l]
Gespenst	gəˈʃpɛnst	/s/ → [θ], /ʃp/ → [ʃ + V + p]
Drache	ˈdʁaxə	/ʁ/ → [ɾ, r, j, l], /dʁ/ → [d + V + ʁ]
Banane	baˈnaːnə	
gelb	gɛlp	
Bagger	ˈbagɐ	
Geburtstag	gəˈbʊɐt͜stak	/t͜s/ → [t͜θ], /k/ → [x]
Jacke	jakə	
streiten	ˈʃtʁaɪtən	/ʁ/ → [ɾ, r, j, l], /ʃtʁ/ → [ʃ + V + t + V + ʁ], /tən/ → [tn]
Ring	ʁɪŋ	/ʁ/ → [ɾ, r, j, l], /ŋ/ → [ng, ŋg], /ŋ/ → [ŋk]
Bild	bɪlt	
Schmetterling	ˈʃmɛtəlɪŋ	/ʃm/ → [ʃ + V + m], /ŋ/ → [ng, ŋg], /ŋ/ → [ŋk]
Bank	baŋk	
Trampolin	ˈtʁampolin	/ʁ/ → [ɾ, r, j, l], /tʁ/ → [t + V + ʁ],
Roller	ˈʁɔlɐ	/ʁ/ → [ɾ, r, j, l]
Buch	buːx	
Spritze	ˈʃpʁɪt͜sə	/ʁ/ → [ɾ, r, j, l], /ʃpʁ/ → [ʃ + V + p + V + ʁ], /t͜s/ → [t͜θ]
Jäger	ˈjɛːgɐ	
Schiff	ʃɪf	
König	ˈkøːnɪç	/ç/ → [ʃ], /ç/ → [k]
Glas	glaːs	/gl/ → [g + V + l], /s/ → [θ]
Auto	ˈaʊto	
blau	blaʊ	/bl/ → [b + V + l],
kaputt	kaˈpʊt	
Schwein	ʃvaɪn	/ʃv/ → [ʃ + V + v]
Sack	zak	/z/ → [ð]
Pflaster	ˈp͜flastɐ	/p͜f/ → [f], /p͜fl/ → [p͜f + V + l], /s/ → [θ]
Trecker / Traktor	ˈtʁɛkɐ / ˈtʁaktɔɐ	/ʁ/ → [ɾ, r, j, l], /tʁ/ → [t + V + ʁ]
Schrank	ʃʁaŋk	/ʁ/ → [ɾ, r, j, l], /ʃʁ/ → [ʃ + V + ʁ]
Apfel	ˈap͜fəl	/p͜fəl/ → [p͜fl]
Zange	ˈt͜saŋə	/t͜s/ → [s], /t͜s/ → [t͜θ], /ŋ/ → [ng], /ŋ/ → [ŋg]
Schere	ˈʃeːʁə	/ʁ/ → [ɾ, r, j, l]
Geschenk	gəˈʃɛŋk	
Fenster	ˈfɛnstɐ	/s/ → [θ]
Brille	ˈbʁɪlə	/ʁ/ → [ɾ, r, j, l], /bʁ/ → [b + V + ʁ]

Wurst	vʊɐst	/s/ → [θ]
Gras	gʁaːs	/ʁ/ → [ɾ, r, j, l], /gʁ/ → [g + V + ʁ], /s/ → [θ]
Schlange	ˈʃlaŋə	/ʃl/ → [ʃ + V + l], /ŋ/ → [ng, ŋg]
Hexe	ˈhɛksə	/s/ → [θ]
Frosch	fʁɔʃ	/ʁ/ → [ɾ, r, j, l], /fʁ/ → [f + V + ʁ]
quak	kvaːk	/kv/ → [k + V +v]
Zwerg	t͜svɛːɐk	/t͜s/ → [s], /t͜s/ → [t͜θ], /t͜sv/ → [t͜s + V +v], /k/ → [ç]
Pferd	p͜fɛɐt	/p͜f/ → [f]
Gießkanne	ˈgiːskanə	/s/ → [θ]
Wippe	ˈvɪpə	
Kleid	klaɪt	/kl/ → [k + V + l]
Nuss	nʊs	/s/ → [θ]
Giraffe	giˈʁafə	/ʁ/ → [ɾ, r, j, l]
Korb	kɔɐp	
Schlüssel	ˈʃlʏsəl	/ʃl/ → [ʃ + V + l], /s/ → [θ], /səl/ → [sl]
Dusche	ˈduːʃə	
Gitarre	giˈtaʁə	/ʁ/ → [ɾ, r, j, l]

Note: CLT, blank cells: item had to be pronounced as in Standard German.

Appendix T: Possible phonetic and dialectal/accentual realisations of the TPA-items scored as correct

Items TPA	Standard Turkish	Accepted phonetic/dialectal/accentual variations
Peynir	pɛjˈniɾ	/ɾ/ → [ɾ̥], /ɾ/ → [ʁ], /ɾ/ → [ɐ]
Güneş	ɟyˈnɛʃ	/ɟ/ → [g]
Bavul	bʌˈʋuɫ	/ʋ/ → [v], /ɫ/ → [l]
Çocuk	t͜ʃoˈd͜ʒuk	/k/ → [x]
Dört	dœɾt	/ɾ/ → [ɾ̥]
Cami	d͜ʒʌːˈmi	
Kalp	kʌlp	/k/ → [g]
Kahve	kʌhˈvɛ	/k/ → [g], /h/ → [x], /h/ → Ø + [ʌː]
Lamba	lʌmˈbʌ	
Nezle	nɛzˈlɛ	/z/ → [ð]
Yağmur	jʌːˈmuɾ	/ɾ/ → [ɾ̥], /ɾ/ → [ʁ], /ɾ/ → [ɐ]
Ev	ɛʋ	/ʋ/ → [v], /ʋ/ → [f]
Dişler	diʃˈlɛɾ	/ɾ/ → [ɾ̥], /ɾ/ → [ʁ], /ɾ/ → [ɐ]
Salyangoz	sʌɫjʌnˈgoz	/s/ → [θ], /ɫ/ → [l], /z/ → [ð]
Mavi	mʌˈvi	
Renk	ɾɛnc	/ɾ/ → [ʁ], /n/ → [ŋ], /c/ → [k]
Siyah	siˈjʌh	/s/ → [θ], /h/ → [x]
Yeşil	jɛˈʃil	
Süt	syt	/s/ → [θ]
Fil	fil	
Garaj	gʌˈɾʌʒ	/g/ → [k], /ɾ/ → [ʁ], /ʒ/ → [ʃ]
Doktor	dokˈtoɾ	/d/ → [t], /ɾ/ → [ɾ̥], /ɾ/ → [ʁ], /ɾ/ → [ɐ]
Itfaiye	itfʌiˈjɛ	
Jöle	ʒœˈlɛ	
Şeftali	ʃɛftʌˈli	
Şişman	ʃiʃˈmʌn	
Zayıf	zʌˈjɯf	/z/ → [ð]
Yorgan	joɾˈgʌn	/ɾ/ → [ʁ], /ɾ/ → [ɐ]
Çay	t͜ʃʌj	/ʌj/ → [aɪ]
Karpuz	kʌɾˈpuz	/k/ → [g], /z/ → [ð]
Zürafa	zyɾʌˈfʌ	/z/ → [ð], /ɾ/ → [ʁ]
Pirinç	piˈɾint͜ʃ	/ɾ/ → [ʁ]
Park	pʌɾc	/c/ → [k], /ɾ/ → [ʁ], /ɾ/ → [ɐ]
Salıncak	sʌɫɯnˈd͜ʒʌk	/s/ → [θ], /ɫ/ → [l], /k/ → [x]
Sandalye	sʌndʌɫˈjɛ	/s/ → [θ], /ɫ/ → [l]
Vagon	vʌˈgon	
Eldiven	ɛldiˈvɛn	
Üçgen	yt͜ʃˈɟɛn	/ɟ/ → [g]
Hediye	hɛdiˈjɛ	

Items TPA	Standard Turkish	Accepted phonetic/dialectal/accentual variations
Ceket	d͜ʒɛˈcɛt	/c/ → [k]
Tavuk	tʌˈʊuk	/t/ → [d], /ʊ/ → [v], /k/ → [x]
vak vak	vʌk vʌk	/k/ → [x]
Çatal	t͜ʃʌˈtʌɫ	/ɫ/ → [l]
Kuş	kuʃ	/k/ → [g]
Iki	iˈci	/c/ → [k]
Ağaç	ʌ:t͜ʃ	
Masa	mʌˈsʌ	/s/ → [θ]
Kurbağa	kuɾˈbʌ:	/k/ → [g], /ɾ/ → [ʁ], /ɾ/ → [ɐ]
Pasta	pʌsˈtʌ	/s/ → [θ]
Gemi	ɟɛˈmi	/ɟ/ → [g]
Rüzgar	ɾyzˈɟʌɾ	/ɾ/ → [ʁ], /z/ → [ð], /ɟ/ → [g], /ɾ/ → [ɾ̥], /ɾ/ → [ɐ]
Anahtar	ʌnʌhˈtʌɾ	/h/ → [x], /ɾ/ → [ɾ̥], /ɾ/ → [ɐ]
Balık	bʌˈɫɯk	/ɫ/ → [l], /k/ → [x]
Havuç	hʌˈʊut͜ʃ	/ʊ/ → [v]
Köpek	cœˈpɛc	/c/ → [k]
hav hav	hʌʊ hʌʊ	/ʊ/ → [v]
Zeytin	zɛjˈtin	/z/ → [ð], /ɛj/ → [ɛɪ]
Gazete	gʌzɛˈtɛ	/g/ → [k], /z/ → [ð]
Ekmek	ɛcˈmɛc	/c/ → [k]
Şapka	ʃʌpˈkʌ	
Gözlük	ɟœzˈlyc	/ɟ/ → [g], /z/ → [ð], /c/ → [k]
Makas	mʌˈkʌs	/s/ → [θ]
Üzüm	yˈzym	/z/ → [ð]
Çiftlik	t͜ʃiftˈlic	/c/ → [k]
Bacak	bʌˈd͜ʒʌk	/k/ → [x]
Jimnastik	ʒimˈnʌstic	/s/ → [θ], /c/ → [k]
Fare	fʌˈɾɛ	/ɾ/ → [ʁ]
Reçel	ɾɛˈt͜ʃɛl	/ɾ/ → [ʁ]
Kitap	citˈʌp	/c/ → [k]
Ejderha	ɛʒdɛɾˈhʌ	/ɾ/ → [ʁ]

Note: CLT, blank cells: item had to be pronounced as in Standard Turkish.

Appendix U: Definitions of phonological patterns and CLT referred to in this thesis

Phonol. pattern / CLT	**Definition**	**Examples**
Affrication	Realising a fricative as an affricate by keeping the same fricative	/ʃɪf/ → [t͜ʃɪf] /zʌˈjɯf/ → [d͜zʌˈjɯf]
Allophonic use of sibilants	Substituting the alveolar, postalveolar and/or palatal sibilants by the same sound	/ˈtaʃə/, /ˈtasə/ → [ˈtaɬə] /zʌˈjɯf/ → [ɮʌˈjɯf]
Assimilation	Adjusting the realisation of a phoneme towards a sound further at the front or back of a word	/kʁokoˈdiːl/ → [kʁokoˈgiːl] /ciˈtʌp/ → [ciˈpʌp]
Backing /s, z, t͜s/	Realising alveolar fricatives (also as part of affricates) as postalveolar or palatal fricatives	/ˈt͜seːbʁa/ → [ˈt͜ʃeːbʁa] /zyrʌˈfʌ/ → [ʒyrʌˈfʌ]
Backing /ʃ, ʒ, t͜ʃ, d͜ʒ/	Realising postalveolar fricatives (also as part of affricates) as palatal fricatives	/ʃɪf/ → [çɪf] /t͜ʃʌˈtʌɫ/ → [t͜çʌˈtʌɫ]
Backing /t, d, n/	Realising alveolar stops/nasals as palatal or velar stops/nasals	/ˈtasə/ → [ˈkasə] /hɛdiˈjɛ/ → [hɛgiˈjɛ]
Biphonemic realisation	Realising one phoneme as two by sounds that combine at least two features with the original phoneme	/ˈjʊŋə/ → [ˈjʊŋgə] /ʁɪŋ/ → [ˈʁɪngə]
Contact assimilation	Adjusting the first member of the CCs /tʁ, dʁ/ to the place of articulation of the second member of the CC	/t͜siˈtʁoːnə/ → [t͜siˈkʁoːnə] /ˈdʁaxə/ → [ˈgʁaxə]
Deaffrication	Reducing an affricate by its stop and realising the fricative only	/tɔp͜f/ → [tɔf] /t͜ʃoˈd͜ʒuk/ → [ʃoˈʒuk]
Denasalisation	Realising a nasal as its non-nasal counterpart	/maˈʁiːn kɛːfɐ/ → [baˈʁiːn kɛːfɐ] /ʌnʌhˈtʌɾ/ → [ʌlʌhˈtʌɾ]
Devoicing	Realising a voiced phoneme as its devoiced counterpart	/ˈhaːzə/ → [ˈhaːsə] /d͜ʒʌːˈmi/ → [t͜ʃʌːˈmi]
Epenthesis of a schwa in unusual positions	Epenthesising a schwa into word positions other than between member of a consonant cluster	/gʁaːs/ → [ˈgʁaːsə] /yt͜ʃˈɟɛn/ → [yt͜ʃəˈɟɛn]
Final consonant cluster reduction	Deleting at least one but not all components of a word- or syllable-final consonant cluster	/hʊnt/ → [hʊn] /t͜ʃiftˈlic/ → [t͜ʃifˈlic]
Final consonant deletion	Deletion of word-final single consonants (i.e., not part of a consonant cluster)	/haʊs/ → [haʊ] /ɛcˈmɛc/ → [ɛcˈmɛ]
Frication	Realising a stop as a fricative	/moˈtoːʁat/ → [moˈsoːʁat] /pɛjˈniɾ/ → [fɛjˈniɾ]
Fronting /c, ɟ, k, g, ŋ/	Realising palatal/velar stops/nasals as their alveolar counterparts	/ˈjakə/ → [ˈjatə] /ˈɟɛmi/ → [ˈdɛmi]

Phonol. pattern / CLT	Definition	Examples
Fronting /ʃ, ʒ, t͜ʃ, d͜ʒ/	Realising postalveolar fricatives (also as part of affricates) as alveolar fricatives	/ˈtaʃə/ → [ˈtasə] /d͜ʒʌːˈmi/ → [dzʌːˈmi]
Gliding of /ɾ/	Realising /ɾ/ as the glides [j, w]	/ɾɛˈt͜ʃɛl/ → [jɛˈt͜ʃɛl] /zyɾʌˈfʌ/ → [ʒywʌˈfʌ]
Glottal substitution of /ʁ/	Realising /ʁ/ as the fricative [h] or glottal stop [ʔ]	/ʁoːt/ → [hoːt] /ˈʃeːʁə/ → [ˈʃeːʔə]
Incorrect stress pattern	Applying the wrong stress pattern to a word	/ˈteːləfoːn/ → [teləˈfoːn] /gʌzɛˈtɛ/ → [gʌˈzɛtɛ]
Initial consonant cluster reduction	Deleting at least one but not all components of a word- or syllable-initial consonant cluster	/ˈtʁɔməl/ → [ˈtɔməl] /ˈt͜seːbʁa/ → [ˈt͜seːʁa]
Initial consonant deletion	Deleting a consonant in word-initial position	/ˈzɔnə/ → [ˈɔnə] /pɛjˈniɾ/ → [ɛjˈniɾ]
Intrusive consonant	Epenthesising an additional consonant into a word	/ʃʁaŋk/ → [ʃtʁaŋk] /gʌzɛˈtɛ/ → [gʌzɛsˈtɛ]
Labialisation	Substituting any type of consonant by a labial/labio-dental of the same manner	/ˈduːʃə/ → [ˈbuːʃə] /sʌɫɯnˈd͜ʒʌk/ → [sʌɫɯmˈd͜ʒʌk]
Liquid deviation	Realising the liquid /ɾ/ as /l/	/jʌːˈmuɾ/ → [jʌːˈmul] /ɾɛnc/ → [lɛnc]
Metathesis	Changing the position of at least two consonants in a word	/nɛst/ → [nɛts] /ciˈtʌp/ → [ciˈpʌt]
Nasalisation	Realising a non-nasal sound as its nasal counterpart	/ˈgaːbəl/ → [ˈgaːməl] /bʌˈɫɯk/ → [bʌˈnɯk]
Stopping fricatives/affricates	Realising a fricative by a stop of the same/nearest place of articulation and realising only the stop of an affricate	/ˈzɔnə/ → [ˈdɔnə] /fil/ → [pil]
Stopping of other sounds than fricatives and affricates	Realising a non-fricative and non-affricate as a stop	/ˈløːvə/ → [ˈdøːvə] /ɾɛˈt͜ʃɛl/ → [dɛˈt͜ʃɛl]
Syllable-final consonant deletion	Deleting a consonant in syllable-final within-word position	/ˈaɪçhœɐnçən/ → [ˈaɪhœɐnçən] /ɛldiˈvɛn/ → [ɛdiˈvɛn]
Syllable-initial consonant deletion	Deleting a consonant in syllable-initial within-word position	/ˈʃeːʁə/ → [ˈʃeːə] /sʌɫɯnˈd͜ʒʌk/ → [sʌɫɯnˈʌk]
Tapping of /l/	Realising the phone /l/ as the tap /ɾ/	/fil/ → [fiɾ] /ɛldiˈvɛn/ → [ɛɾdiˈvɛn]
Unusual fronting	Realising a more posterior phone as an anterior phone of the same manner but not following the general fronting patterns	/ˈdʁaxə/ → [ˈdʁasə] /siˈjʌh/ → [siˈlʌh]

Phonol. pattern / CLT	Definition	Examples
Vocalisation of /l/	Realising the phone /l/ as the glide /j/ (making it sound more like a vowel)	/ˈʁɔlɐ/ → [ˈʁɔjɐ] /sʌɫɯnˈd͜ʒʌk/ → [sʌjɯnˈd͜ʒʌk]
Vocalisation of /ɾ/	Realising the /ɾ/ as a vowel in line with the German phonotactic rule to substitute /ʁ/ as [ɐ] in word-final positions	/pɛjˈniɾ/ → [pɛjˈniɐ] /dœɾt/ → [dœɐt]
Voicing	Realising a devoiced phoneme as its voiced counterpart	/giˈʁafə/ → [giˈʁavə] /ciˈtʌp/ → [ciˈdʌp]
Vowel epenthesis in CC	Epenthesising a vowel (usually the schwa) between the components of a consonant cluster	/ˈbluːmə/ → [ˈbəˈluːmə] /ˈt͜seːbʁa/ → [ˈt͜seːbəʁa]
Weak syllable deletion	Deletion of an unstressed/weak syllable in any position of the word	/gəˈbʊɐt͜stak/ → [ˈbʊɐt͜stak] /ʌnʌhˈtʌɾ/ → [ʌnˈtʌɾ]
Word-final devoicing	Realising voiced fricatives in word-final position as their devoiced counterparts	/muz/ → [mus] /ɛʋ/ → [ɛf]

Appendix V: Correlations between the amount of input and output within and across both languages at t1

		Output	
		German	Turkish
Input	German	.842***	-.356**
	Turkish	-.213	.813***

Note: very strong correlation, weak correlation, *** p < .001, ** p < .01

Appendix W: Phonological profiles and categorisation of children who only participated at t1 (N = 41)

P	Sex	Age	Phonological patterns		Infrequent variants		Category
			German	Turkish	German	Turkish	
42	M	3;2	T	T	+	+	T
63	F	3;7	T	T	+	+	T
4	F	3;11	T	T	+	+	T
21	F	3;11	T	T	+	+	T
20	F	4;4	T	T	+	+	T
77	F	4;5	T	T	+	+	T
24	M	4;6	T	T	+	+	T
51	M	4;6	T	T	+	+	T
2	F	4;7	T	T	+	+	T
57	F	4;7	T	T	+	+	T
76	F	4;7	T	T	+	+	T
41	F	4;8	T	T	+	+	T
70	F	4;8	T	T	+	+	T
23	F	4;9	T	T	+	+	T
14	F	4;10	T	T	+	+	T
66	F	4;10	T	T	+	+	T
25	M	4;11	T	T	+	+	T
47	F	5;0	T	T	+	+	T
69	M	5;1	T	T	+	+	T
12	M	3;6	T	Dev	-	-	A
75	M	3;9	T	T	-	-	A
43	F	3;10	Del	Del	+	+	A
48	M	4;4	Del	Del	+	-	A
81	M	4;4	Del	Del	+	+	A
78	F	4;10	Del	Del	+	-	A
79	M	5;3	Del	T	-	-	A
83	F	3;0	T	Del	+	+	N
82	M	3;1	Dev	T	+	+	N
72	F	3;2	T	T	-	+	N
80	M	3;4	Del	T	+	+	N
84	M	3;4	Dev	T	-	+	N
15	M	3;5	Dev	T	+	+	N
1	M	3;11	Dev	T	+	+	N
73	F	4;0	T	Del	+	+	N
74	F	4;0	T	T	-	+	N
40	F	4;1	Dev	T	+	+	N
19	F	4;4	Del	T	+	+	N
71	M	4;7	Del	T	-	+	N
44	M	4;9	T	Dev	+	-	N
46	M	4;11	T	Del	+	-	N
67	F	4;11	Del	T	+	+	N

Note: **P**: Participant, German, Turkish, **T**: typical, **Dev**: deviant, **Del**: delayed, **+**: typical performance (≤ 1.25 SD above the age group mean) **-**: weak performance (> 1.25 SD above the age group mean), **M**: male, **F**: female, **T**: categorised as typical, **A**: categorised as atypical, **N**: no category

Appendix X: Phonological profiles and categorisation of children who were retested at t2 (N = 43)

P	Sex	Age t1	Age t2	Patterns t1		Patterns t2		InfVar t1		InfVar t2		NWR	Cons.	PI	Cat t1	Cat t2
				Ger	Tur	Ger	Tur	Ger	Tur	Ger	Tur					
10	F	3;1	4;5	T	T	T	T	+	+	+	+	+	+	+	T	T
59	F	3;3	4;3	T	T	T	T	+	+	+	+	+	+	+	T	T
22	F	3;7	4;10	T	T	T	T	+	+	+	+	+	+	+	T	T
31	F	3;9	5;1	T	T	T	T	+	+	+	+	+	+	+	T	T
32	F	3;10	5;0	T	T	T	T	+	+	+	+	+	+	+	T	T
3	F	3;11	5;4	T	T	T	T	+	+	+	+	+	+	+	T	T
28	F	4;1	5;3	T	T	T	T	+	+	+	+	+	+	+	T	T
37	M	4;2	5;5	T	T	T	T	+	+	+	+	+	+	+	T	T
18	F	4;5	5;7	T	T	T	T	+	+	+	+	+	+	+	T	T
27	M	4;5	5;7	T	T	T	T	+	+	+	+	+	+	+	T	T
30	M	4;6	5;9	T	T	T	T	+	+	+	+	+	+	+	T	T
53	M	4;8	5;8	T	T	T	T	+	+	+	+	+	+	+	T	T
58	M	4;10	5;10	T	T	T	T	+	+	+	+	+	+	+	T	T
39	F	4;11	6;1	T	T	T	T	+	+	+	+	+	+	+	T	T
55	M	5;1	6;2	T	T	T	T	+	+	+	+	+	+	+	T	T
8	M	3;1	4;6	T	T	T	Del	+	+	+	+	+	+	+	T	N
17	F	3;6	4;6	T	T	T	T	+	+	-	+	+	+	+	T	N
65	M	3;7	4;9	T	T	T	Del	+	+	+	+	+	+	+	T	N
26	F	3;9	5;0	T	T	T	T	+	+	+	-	-	+	+	T	N
7	F	3;10	5;3	T	T	T	Del	+	+	+	+	+	+	+	T	N
45	M	3;10	4;10	T	T	Del	T	+	+	+	+	+	+	+	T	N
56	F	3;11	4;11	T	T	Del	T	+	+	+	+	+	+	+	T	N
50	F	4;6	5;6	T	T	T	Del	+	+	+	-	+	+	-	T	N
6	F	4;10	6;3	T	T	T	Del	+	+	+	+	+	+	+	T	N
16	M	5;1	6;3	T	T	T	Dev	+	+	+	-	+	+	+	T	N
62	F	5;1	6;1	T	T	T	T	+	+	-	+	+	+	+	T	N
5	F	4;1	5;6	Del	Del	T	T	-	+	+	+	+	+	+	A	T
60	F	4;11	5;11	Del	Del	T	T	+	+	+	+	+	+	+	A	T
49	M	4;5	5;5	Dev	Dev	Dev	Dev	+	+	+	+	+	+	-	A	A
36	M	4;7	5;9	Dev	Dev	Dev	Dev	-	+	+	+	+	+	-	A	A
11	F	3;0	4;5	T	Dev	T	T	+	+	+	+	+	+	+	N	T
54	F	3;1	4;1	Del	T	T	T	+	+	+	+	+	+	+	N	T
29	F	3;7	4;10	T	T	T	T	-	+	+	+	+	+	+	N	T
13	M	3;9	5;1	Del	T	T	T	+	+	+	+	+	+	+	N	T
52	M	3;11	4;11	Del	T	T	T	+	+	+	+	+	+	+	N	T
64	M	3;9	4;10	Dev	T	T	T	+	+	-	-	-	-	-	N	A
34	F	4;0	5;2	Dev	T	Del	T	+	-	-	-	+/-	+/-	+	N	A
33	M	4;7	5;10	Dev	T	Del	T	-	+	-	-	-	-	+	N	A
35	F	4;1	5;4	Dev	T	Dev	T	+	+	+	+	-	+/-	+	N	N
38	M	4;5	5;7	Del	T	Dev	T	+	+	+	+	-	+	+	N	N
68	F	4;6	5;7	T	Del	T	Del	+	+	+	+	+/-	+	+	N	N
61	M	4;9	5;9	T	T	T	Del	+	-	+	+	+	+	+	N	N
9	M	5;1	6;5	T	Del	T	Del	+	+	+	+	+	-	+	N	N

Note: **P**: Participant, **Cons.**: consistency task, **PI**: phone imitation, **Cat**: Category, **Ger**: German, **Tur**: Turkish, **T**: typical, **Dev**: deviant, **Del**: delayed, **+**: typical performance (≤ 1.25 SD above age group mean for InfVar and below age group mean for NWR, Cons and PI), **-**: weak performance (> 1.25 SD above age group mean for InfVar and below age group mean for NWR, Cons and PI), **+/-**: borderline performance (> 1 SD but < 1.25 SD below age group mean), **M**: male, **F**: female, **T**: typical, **A**: atypical, **N**: no category

Appendix Y: Results of the hierarchical multiple regression analyses of psycholinguistic skills predicting single-word naming skills at t2

Multiple regression for the predictive measures of PhonVar in German at t2

Predictors	Entered first		Entered last		
	R^2 change	*Sig. F change*	*R^2 change*	*Sig. F change*	*B*
NWR	.405	< .001	.098	.013	-3.855
Consistency	.207	.002	.000	.988	0.012
Phone imitation	.255	.001	.029	.029	-1.702
R^2 final model (incl. all predictors)			**.435**	**< .001**	

Multiple regression for the predictive measures of PhonVar in Turkish at t2

Predictors	Entered first		Entered last		
	R^2 change	*Sig. F change*	*R^2 change*	*Sig. F change*	*B*
NWR	.494	< .001	.113	< .001	-3.761
Consistency	.175	.005	.012	.249	0.624
Phone imitation	.515	< .001	.141	< .001	-3.382
R^2 final model (incl. all predictors)			**.654**	**< .001**	

Multiple regression for the predictive measures of InfVar in German at t2

Predictors	Entered first		Entered last		
	R^2 change	*Sig. F change*	*R^2 change*	*Sig. F change*	*B*
NWR	.348	< .001	.053	.054	-1.597
Consistency	.420	< .001	.092	.012	-1.069
Phone imitation	.031	.256	.029	.149	0.953
R^2 final model (incl. all predictors)			**.478**	**< .001**	

Multiple regression for the predictive measures of InfVar in Turkish at t2

Predictors	Entered first		Entered last		
	R^2 change	*Sig. F change*	*R^2 change*	*Sig. F change*	*B*
NWR	.533	< .001	.116	.003	-2.422
Consistency	.386	< .001	.017	.232	-0.469
Phone imitation	.190	.003	.001	.755	-0.193
R^2 final model (incl. all predictors)			**.550**	**< .001**	

Multiple regression for the predictor measures of PCC-A-scores in German at t2

Predictors	Entered first		Entered last		
	R^2 change	*Sig. F change*	*R^2 change*	*Sig. F change*	*B*
NWR	.366	< .001	.054	.065	0.924
Consistency	.221	.001	.003	.637	0.117
Phone imitation	.277	< .001	.050	.074	0.718
R^2 final model (incl. all predictors)			**.418**	**< .001**	

Multiple regression for the predictor measures of PCC-A-scores in Turkish at t2

Predictors	Entered first		Entered last		
	R^2 change	*Sig. F change*	*R^2 change*	*Sig. F change*	*B*
NWR	.561	< .001	.161	< .001	1.836
Consistency	.204	.002	.014	.211	-0.271
Phone imitation	.463	< .001	.089	.002	1.097
R^2 final model (incl. all predictors)			**.669**	**< .001**	