REVISTA DA **ABRALIN**

LITERATURE REVISION

Phonetic transcription of spontaneous children's speech with the aid of software: a systematic review

Débora Tomazi Moreira CAUMO
Universidade Federal do Rio Grande do Sul (UFRGS)

Márcio Pezzini FRANÇA
Universidade Federal do Rio Grande do Sul (UFRGS)

Clécio Homrich da SILVA
Universidade Federal do Rio Grande do Sul (UFRGS)

OPEN ACCESS

EDITED BY
- Raquel Freitag (UFS)

REVIEWED BY
- Andressa Toni (USP)
- Márcia Cristina do Carmo (UEPG)

ABOUT THE AUTHORS
- Débora Tomazi Moreira Caumo
Conceptualization, Methodology, Supervision, and Writing of the article.
- Márcio Pezzini França
Conceptualization, Methodology, Supervision, and Writing of the article.
- Clécio Homrich da Silva
Conceptualization, Methodology, Supervision, and Writing of the article.

DATES
- Received: 29/06/2022
- Accepted: 22/12/2022
- Published: 17/01/2022

HOW TO CITE
Caumo, D. T. M.; França, M. P., Silva, C. S. (2022), Phonetic transcription of spontaneous children's speech with the aid of software: a systematic review. *Revista da Abralin*, v. 21, n. 1, p. 1-22, 2022.

ABSTRACT

The aim of the study was to identify, synthesize and classify the software currently available that can help in the task of phonetic transcription of the spontaneous speech of pre-school children to evaluate the development of children's language. A systematic review was performed for articles published, for the 10-year period (June 2010 to June 2020), without restrictions as to location and language, using the Cochrane, Pubmed and Web of Science databases. The terms used in the search strategies were "phonological", "phonetic", "transcription", "computer" and "software". The studies were selected by two independent reviewers using pre-defined search strategies. In the initial search, after the exclusion of duplicates, 534 articles were found. By reading their titles and abstracts, 46 articles related to the theme were left, which were then read in full. After reading, 24 articles were included in the study. The results revealed a total of seven software available for the phonetic transcription of spontaneous speech from preschoolers used for different analyses: LENA and Timestamper (for babbling and pre-linguistic vocalizations), ELAN (for gestural communication, extralinguistic elements and the

situational context), Phon (for phonetic and phonological analyses), CLAN and SALT (for morphosyntactic, grammatical and semantic aspects) and Praat (for acoustic measurements). Through this systematic review, it can be concluded that there are advantages to using software for phonetic transcription, sample storage, and child language analysis, especially concerning standardization and reliability for spontaneous speech samples. Phonetic transcription still relies on the ability and subjectivity of a human transcriber. The tools found in the software provide support to facilitate using phonetic symbols, audio segmentation and pairing to writing, and analysis of speech data.

RESUMO

O objetivo do estudo foi identificar, sintetizar e classificar os softwares atualmente disponíveis que podem auxiliar na tarefa de transcrição fonética da fala espontânea de pré-escolares, para avaliar o desenvolvimento da linguagem infantil. Foi realizada uma revisão sistemática de artigos publicados, no período de 10 anos (de junho de 2010 a junho de 2020), sem restrições quanto à localização e idioma, utilizando as bases de dados Cochrane, Pubmed e Web of Science. Os termos utilizados nas estratégias de busca foram "fonológico", "fonético", "transcrição", "computador" e "software". Os estudos foram selecionados por dois revisores independentes usando estratégias de busca pré-definidas. Na busca inicial, após a exclusão de duplicatas, foram encontrados 534 artigos. Com a leitura de seus títulos e resumos, restaram 46 artigos relacionados ao tema, que foram lidos na íntegra. Após a leitura, 24 artigos foram incluídos no estudo. Os resultados revelaram um total de sete softwares disponíveis para auxiliar a transcrição fonética da fala espontânea de pré-escolares utilizados para diferentes análises: LENA e Timestamper (para balbucios e vocalizações pré-linguísticas), ELAN (para comunicação gestual, elementos extralinguísticos e contexto situacional), Phon (para análises fonéticas e fonológicas), CLAN e SALT (para aspectos morfossintáticos, gramaticais e semânticos) e Praat (para medidas acústicas). Por meio desta revisão sistemática, pode-se concluir que há vantagens no uso de software para transcrição fonética, armazenamento de amostras e análise de linguagem infantil, principalmente no que diz respeito à padronização e confiabilidade para amostras de fala espontânea. A transcrição fonética ainda depende de um transcritor humano. As ferramentas encontradas nos softwares fornecem suporte para facilitar o uso dos símbolos fonéticos, segmentação e pareamento de áudio para escrita e análises de dados de fala.

KEYWORDS
Phonetics, Child Language, Software, Language Development.

PALAVRAS-CHAVE
Fonética. Linguagem Infantil. Software. Desenvolvimento da Linguagem.

SUMMARY FOR NON-SPECIALISTS
Phonetic transcription is a method of describing speech. Speech sounds are represented by special graphic symbols. To assess speech and language in children, it is often necessary to transcribe recordings of their spontaneous speech. Assessing speech in a natural context is especially important to consider the skill set, communicative context and sociocultural aspects. In this systematic review, all studies with preschool-age children were included to identify, synthesize and classify existing software that can help in the phonetic transcription of spontaneous speech of children in this age group.

Introduction

To assess speech and language in children, it is often necessary to transcribe recordings of their spontaneous speech. Assessing speech in a natural context is especially important to consider skill set, communicative context, and socio-cultural aspects. Phonetic transcription is a method of describing spontaneous speech. Speech sounds are represented by special graphic symbols. Usually based on the International Phonetic Alphabet, the IPA is a system of standard notation for the phonetic representation of all languages in the world that was developed by phoneticians from the International Phonetic Association. The latest version of the IPA Alphabet was published in 2015, and IPA charts are re-issued annually (INTERNATIONAL PHONETIC ASSOCIATION, 2021).

The term "spontaneous" describes situations in which the speaker's verbal productions are not fully induced and evoked, therefore allowing participants creative freedom (GENEST; MASSON, 2017). In some clinical situations, when the verification of language skills is semi-guided, spontaneous speech production can be part of the assessment. There are other situations in which spontaneous speech can occur in which we find the terms "ecological" or "natural" as research data. The terms "ecological" and "natural" refer to the situations and contexts of data collection, such as the environments in which the child builds relationships (e.g. at home and school). These environments provide the opportunity to collect speech samples during natural, routine activities.

Phonetic transcriptions are examiner-dependent, and reproducibility is used as a measure of reliability. Intra-examiner reliability refers to the consistency of measurements taken under the

same conditions of assessment, at two or more different times, by the same examiner. Inter-examiner reliability, on the other hand, is linked to the consistency of measurements performed by two or more examiners (VENTURINI et al., 2006).

Traditionally, phonetic transcriptions are manually annotated and collected by specialists to guarantee quality. However, in many cases, this can limit the reliability of the information (ALVAREZ et. al., 2015). Even with the advancement of technology, software capable of performing speech recognition for automatic subtitles they are error prone and require revision. This is an even greater challenge for speech with deviations. Speech of preschool-age children usually present phonological errors and other types of language problems (e.g. morphological and syntactic errors). In such situations, analyses performed by specialists trained in transcription of child speech are preferred (SABRI, 2018).

In Brazil, children are considered to be of preschool age until they are 6 years old. They begin elementary school at that time. In the United States of America, children are considered preschoolers until they reach the age of five (FREITAS; SHELTON, 2005).

Phonetic transcription is a useful means for language assessments of preschool children through a sample of spontaneous speech. But it can be laborious, time-consuming and costly. Children are developing speech skills, which can make speech even more difficult to transcribe. Technology is constantly advancing and can help in this task.

The main objective of this study was to find software that assists in performing phonetic transcription of children's speech. The specific objectives were to investigate the use of software in research with children's spontaneous speech in a decade of publications of scientific articles, to identify, synthesize and classify existing software that can help in the task of phonetic transcription.

1. Method

This systematic review was carried out using the Cochrane Central Register, PubMed, and Web of Science databases, and registered on the PROSPERO platform - International Prospective Register of Systematic Reviews - online database of systematic review protocols on health-related topics, under CRD number 42020191698 on 07/26/2020. On this date, there were still no records of other reviews in progress or published on this topic. This certified that our research is unprecedented and relevant. The Appendix 1 presents more details of the methodology applied and the registration of this project on the PROSPERO platform. The Mesh (Medical Subject Heading) terms “phonological”, “phonetic”, “transcription”, “computer” and “software” were used for each of the selected databases (Cochrane Central Register, PubMed and Web of Science) with search strategies formulated by a librarian (see Appendix 1).

The articles were first selected by two independent reviewers according to the pre-defined search strategies. Duplicate articles in the databases were excluded. After reading the titles and abstracts, those that were not related to the research topic were excluded. The resulting files were saved in the Zotero reference manager and read in their complete format for analysis using inclusion, exclusion, and final selection criteria.

The steps for selecting and analysing the articles, as well as presenting the results, were performed according to the PRISMA protocol - Preferred Reporting Items for Systematic Reviews and Meta-Analyses- (MOHER et al., 2009).

Inclusion criteria

- Previously published in journals, specialist magazines, or even indexed in the referred databases for the 10-year period from June 2010 to June 2020, with no restrictions as to location and language.

- Studies that used software to assist phonetic transcription of the spontaneous speech with audio and / or video recording for phonetic transcription.

- Studies that included children of preschool age, admitting studies that contain other age groups in addition to preschool children.

- Exclusion criteria

- Publications that used software for a purpose other than phonetic transcription of spontaneous speech recordings.

We found 645 records in the initial search of the three databases and duplicates were excluded (111), leaving 534 records. The search resulted in the following records: Cochrane Central Register, 1; PubMed, 171 and Web of Science, 473. After the exclusion of articles with unrelated titles and abstracts, 46 records remained, which were read in full analysis. The articles were excluded at this stage (22) because they did not fit the criteria of the children's age range (non-inclusion of preschoolers) and the use of software that does not allow the assessment of spontaneous speech. In this way, 24 studies were included for the qualitative synthesis, as presented in the flow diagram (Figure 1).

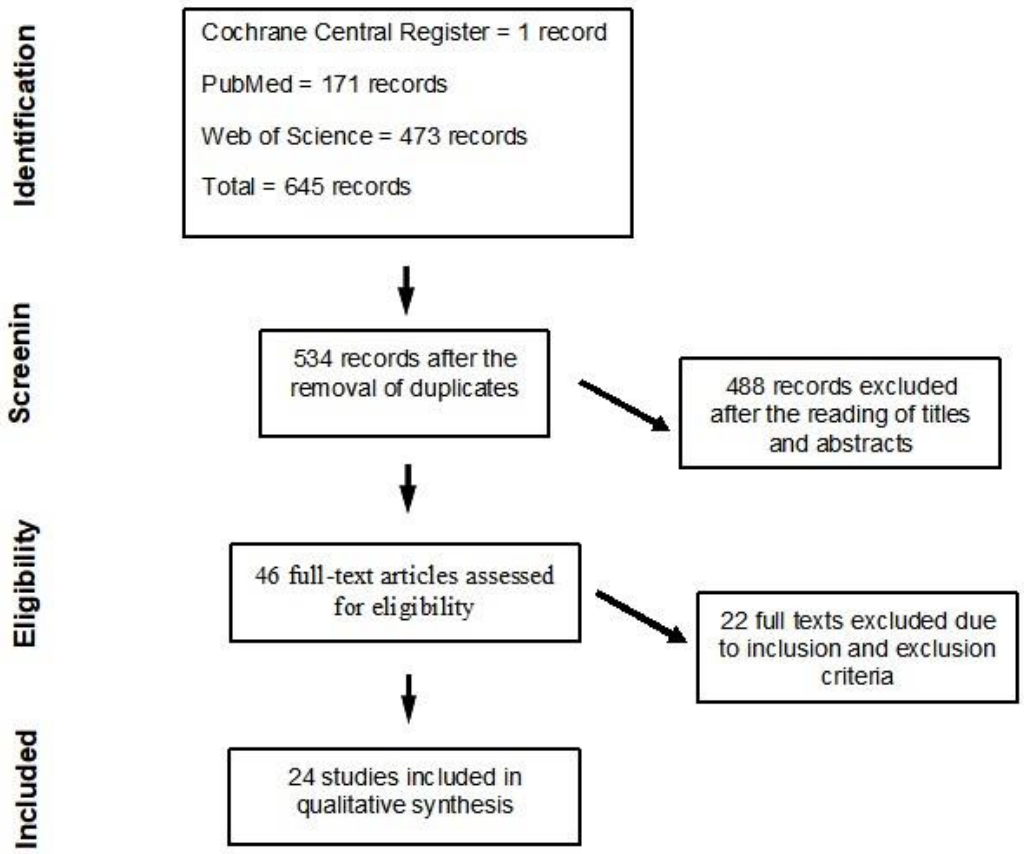

FIGURE 1 – Study flow diagram of the search and selection process.
Source: authors.

2. Results

The summaries of the 24 articles selected for this systematic review are presented in Table 1, with information regarding the author(s), years of publication, objectives, methods, participant ages, sample sizes, and software.

Author/ Year	Objective	Method	Sample		Software
			Age (a)	Size and sex	
SABRI; FABIANO-SMITH, 2018	To observe the interaction between two languages and the effect of cochlear implants (CIs) on the acquisition of two phonetic systems.	A longitudinal study that examined the phonological development of an Arabic-English bilingual child with bilateral CIs.	3.6-4.0 years	1 F	Phon
BURGESS ET AL., 2013	To characterize and compare the home and school linguistic environments of preschool-aged children with Autism Spectrum Disorders (ASDs).	A longitudinal study in which naturalistic language samples from the children were recorded at three-month intervals for one year.	35-67 months	10 F and M	LENA, SALT
FLIPSEN; KANGAS, 2014	To document the mean length of utterance (MLU) in children using CIs due to profound hearing loss.	A longitudinal study in which samples were obtained by postgraduate students trained to interact with and record the children in audiometric booths.	3 years	10 F and M	SALT
XU et al., 2014	To analyze early language development of language in a comparison between an automated method and a conventional method of manual transcription.	A case study using an automated analysis to study naturalistic recordings of child phonetic production	8-48 months	106 F and M	LENA
D'APICE et al., 2019	To verify the contribution of parenting to the language, cognition, and behavior of children.	A cross-sectional study including a large sample of children and their families studied in their home environment over three days.	2.03-3.99 years	107 F and M	LENA, CLAN
COLLETTA et al., 2010	To investigate changes in the development of discourse that occur in language and gestures.	A cross-sectional study comparing three age-group samples (6 years, 10 years, and adults) and including 84 children and a total of 122 participants.	5.3-10 years	84 F and M	ELAN
RATNER; MACWHINNEY, 2016	To assess the advantages of analyzing speech samples with speech analysis software.	Sample research of a broader study of a large cohort of mothers and babies, to investigate the possible predictors of later language performance by the children, and review the data obtained from a data exchange system.	7-24 months	125 F and M	CLAN
CANAULT, et al., 2016	To examine the accuracy of LENA, a language environment analysis system, in European French.	A comparison between the automated and human methods of audio recording analysis for counting human adult words and child vocalization.	3-48 months	18 F and M sexes	LENA
PEZOLD et al., 2020	To analyze speech samples by comparing tests run by two computer programs.	A study of two prematurely born, preschool-aged children included in the database of a previous study	4.9 years	2 F and M	CLAN, SALT
OLLER et al., 2010	To obtain development measures for early speech.	A comparison between the automated analysis of audio recordings and conventional methods of manual transcription for children presenting typical development, LLE, and ASDs.	16-48 months	232 F and M	LENA
BALCIUNIENE, 2012	To analyze the narrative characteristics of Lithuanian pre-schoolers.	A cross-sectional study with an analysis based on experimental data from child narratives collected at a nursery school.	6-7 years	24 F and M	CLAN
GREENWOOD et al., 2011	To examine the relation between the volume of vocal productions and the spoken vocabulary of children, using automatic speech processing.	A prospective longitudinal, cross-sectional study in which audio recordings were repeatedly collected from home environments over 10 months.	12-20 months	30 F and M	LENA

ROSE; STOEL-GAMMON, 2015	To present an overview of new tools that can be used to deepen our understanding of phonological development and disorders.	A case study with the longitudinal data of one child with typical phonological development and three children with a phonological disorder.	2.0-5.8 years	4 M	Phon, Praat
MOELLER et al., 2010	To examine the impact of late-identified hearing loss on speech and language.	A longitudinal study with a control group and a descriptive analysis of cases involving children with slight to moderate sensorineural hearing loss.	28-41 months	14 F and M	SALT
CHAPARRO-MORENO, REALI; MALDONADO-CARREÑO, 2017	To use reading tasks to investigate the linguistic productions of Colombian preschool children.	A cross-sectional study in which a sample of 13 teachers matched with groups of four children from a private school shared reading sessions. These sessions were recorded and analyzed and the effects of reading an illustrated book with and without words were compared.	43-55 months	52 F and M	CLAN
RASANEN et al., 2019	To assess a system with the recorded samples of children from six different linguistic corpora.	The development of an open-source system that can be adapted to different languages or dialects, with orthographically transcribed speech data in the corpora.	0-32 months	58 F and M	LENA
BREDIN-OJA et al., 2018	To investigate the reliability of an automated language analysis system in comparison with human transcribing.	A case study determining the rate of vocalization in children with different pathologies that involve language impairment.	31-46 months	6 M	LENA
BUSCH et al., 2018	To assess the reliability of LENA in comparison with manual transcription.	A correlation and concordance study between LENA estimates and a manual count of 48 audio recording samples.	2-5 years	6 F and M	LENA, Praat
JULIEN et al., 2019	To examine the level of morphosyntactic development in neglected children.	A cross-sectional study, with a control group, was part of a broader longitudinal study measuring levels of language development and personal characteristics in the home environment.	4 years	170 F and M	SALT
GENEST; MASSON, 2017	To describe methods for transcribing data in natural environments.	A description of the linguistic tools for processing oral data, to present the principles and methods for building a linguistic corpus.	NA	NA	CLAN, Praat
GOMES DE MELO BEZERRA et al., 2016	To contribute to the research of language acquisition through the analysis of natural speech interactions.	A description of ELAN and CLAN software programs and the transcription of language data.	NA	NA	ELAN, CLAN.
CANAULT; LE NORMAND; THAI VAN, 2017	To give an overview of the LENA system, as well as its advantages and disadvantages.	A description of the uses for the software analysis.	NA	NA	LENA, Praat
BYUN; ROSE, 2016	To present Phon software and a series of functions specially designed to facilitate the study of child phonology.	A description of the software tools and uses.	NA	NA	Phon
WILLADSEN et al.,2018	To develop software for codifying infant babbling.	The development of a software program to assist with codifying pre-linguistic real-time vocalizations in nursing infants with cleft palate.	NA	NA	Time-Stamper

(a) M and F (masculine and feminine), CI (cochlear implant), ASDs (autism spectrum disorders), LLE (late language emergence).

(b) NA (not applicable).

TABLE 1 - Synthesis of studies using phonetic transcription software to analyze language development in children: objectives, methods, samples, and software.
Source: authors.

From these 24 studies, we found seven software programs designed to assist phonetic transcription of the spontaneous speech of preschool-aged children. All software was developed in the English language, although some studies were also carried out in other languages. Table 2 shows these software programs, the language of each study, and the cost of access.

Software	**Research language**	**Operational System**	**Access: Free/ Non-free**
LENA	French, American English, British English, Dutch	Not mentioned.	Non-free
TimeStamper	British English, Germanic Scandinavian, Brazilian Portuguese	Windows, Mac, Linux.	Free
ELAN	French, Brazilian Portuguese	Windows, Mac, Linux	Free
Phon	American English, British English, Arabic	Windows, Mac, Linux	Free
CLAN	Spanish, American English, French, Lithuanian, Brazilian Portuguese	Windows, Mac, Linux	Free
SALT	French, American English	Windows, Mac	Non-free
Praat	French, Dutch, American English, British English	Windows, Mac, Linux,	Free

TABLE 2 – Software programs, research languages, operational systems, and cost of access.
Source: authors.

The types of analysis performed by the software were classified into four categories:

- Pre-linguistic communication and prosody - analyses encompassing the number of child vocalizations (NCV), babbling, the counting of child conversational turns, the counting of transcribed syllables, vocal and prosodic aspects (acoustic signals of speech), and linguistic exposure with adult word counts (AWC).

- Phonetics and phonology - inventories of consonants and vowels; syllable forms and words; stress patterns, relational analyses (identifying substitution and/or exclusion errors based on individual phonemes or phonological classes), the Percentage of Consonants Correct (PCC) or Percentage of Vowels Correct (PVC) metric, in addition to precision measures of whole words.

- Morphosyntax, grammar, and semantics - frequency counts, such as the total number of words, the number of different words (NDW), narrative productivity, type of utterance, co-occurrence analysis, abandoned utterances, verbal rate, mean length of utterance (MLU), frequency and lexical diversity, and type/token ratio (TTR - a measure of lexical diversity, in which the number of unique words is divided by the total number of words).

- Extra-linguistic communication, gestures, and context - facial expressions, body posture, actions, handling, information about the environment and surrounding events, and any other information outside of vocal or speech production.

Results revealed a total of seven software being used to aid phonetic transcription of the spontaneous speech of preschoolers, in different types of analysis: LENA and TimeStamper (used for studying babbling and pre-linguistic vocalizations), ELAN (used to analyze gestural communication, with the transcription of extra-linguistic elements and situational contexts), Phon for phonetics and phonology, CLAN and SALT for the analysis of morphosyntax, grammar, and semantics, and Praat for analysis of sound waves and the parameters of frequency, duration, and intensity (see Figure 2).

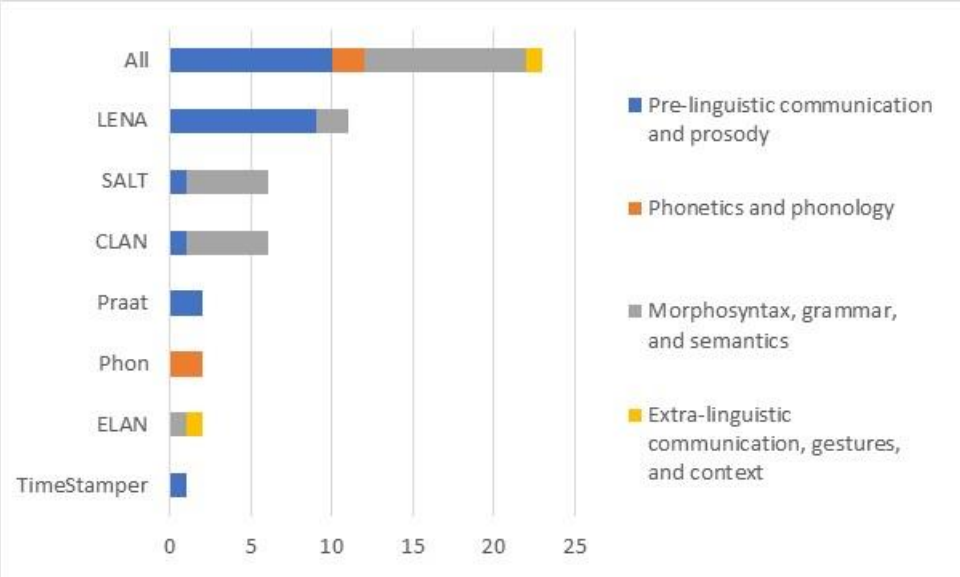

FIGURE 2. Types of analysis applied in language studies using software.
Source: authors.

The full function of these software programs is not restricted to only the features described in the studies. Software information on the operation and applicability for aiding Phonetic Transcription is presented below:

1. LENA: It is a digital audio recording device for transcription. Private access (http://www.lena.org). Provides automated measures of speech heard and produced by the child. The algorithms are trained to identify and differentiate speech from adults, from children, from noise. It does not recognize any words that are being spoken. Since the algorithms are not open source, it is not possible to improve the software or update it. The WCE (Word Count Estimation) module has been enhanced for American English. Although the system can be used with recordings in any language, its accuracy is not necessarily consistent across different populations and this complicates any attempt to compare languages (CANAULT, et al., 2016), (RÄSÄNEN, et al., 2019).

2. TimeStamper: Plays video (mp4) or audio (mp3) files. It can be obtained by contacting ctr-cis@liverpool.ac.uk and will be made available for non-commercial use under the terms of the Apache 2.0 license. TimeStamper assists the transcriber in the evaluation of canonical babbling, to verify whether or not the child was observed in the canonical babbling stage. At the end of the recording, a window appears, asking the encoder to select yes or no for canonical babbling, and lists the syllables presented. Responses are automatically saved in two different locations, making it easy to analysis data and assign a particular recording to specific transcribers. A transcriber can only see the last entry to avoid any influence on final decisions. The canonical babbling index can be calculated automatically, and it is possible to view the syllables presented by the child in a syllabic inventory. At the end of the recording, a window appears asking the encoder to select canonical yes/no, and then lists the syllables it has found. Real-time evaluation reduces time and resource demand, but uploading a new video will result in the loss of all previous data (WILLADSEN et al., 2018).

3. ELAN: Allows you to link up to 4 video files to an annotation document. It uses existing native media structures, such as Windows Media Player, QuickTime or VLC. Free access (https://archive.mpi.nl/tla/elan). Performs linguistic analysis, allows data notes of oral and non-verbal language (gestures). It makes it possible to add an unlimited number of text annotations to audio and/or video recordings aligned with the media time, but it depends on the user's notes (GOMES DE MELO BEZERRA et al., 2016).

4. Phon: Provides a framework for building multimedia databases (mp3, mp4). Free access (https://www.phon.ca). Includes time alignment of multimedia data with transcription; resources with IPA characters (IPA map; phonetic dictionaries for different languages); interface for multiple blind transcriptions and consensus-based transcription validation; systems for automatic data identification (phonetic characteristics; syllabification). Enables data consultation and produces reports. The built-in dictionaries, despite offering automatic phonetic transcription through the insertion of orthographic writing by the user, are not yet available in all languages (ROSE; STOEL-GAMMON, 2015).

5. CLAN: Loads audio and video data (mp3, mp4). Free access (http://childes.talkbank.org). It allows the transcription and aligning it whith audio or video. Additional information (gestures, interpretations) can be documented using conventions to transcribe the different statements. A file transcribed with CLAN software consists of header lines starting with the @ symbol and contains corpus metadata such as language, participants, dates, locations, and video names. In the transcription, each speech production is associated with a "tag" that isolates the audio or video passage in question. A set of transcription conventions must be used for linguistic phenomena that are not uttered by the study participant. The choice and number of lines must be determined by the researcher and/or the professional, according to the aims of the

analysis. CLAN software also offers the means to export data to other software, to generate additional analyses, (GENEST; MASSON, 2017), and accommodates investigations concerning frequency counts and morphosyntax (GOMES DE MELO BEZERRA et al., 2016). Although CLAN can complete analyses in 49 languages, the comparisons in the database and the value list published in the manual are only for English-speaking children. CLAN users can compare samples by exchanging with database systems that the software is part of (PEZOLD et al., 2020).

6. SALT: Most digital audio formats are supported. Private access (http://saltsoftware.com/products/software). Standardizes the process of obtaining, transcribing and analyzing language samples. It contains a transcription editor and a database of normative conversations in English and Spanish that allow the automatic calculation of standard scores to aid clinical judgment. This software has an integrated training program designed to extend its use as a clinical tool to non-clinical users who are not speech therapists. SALT tools include several different types of automatic analysis, such as average length of statement, number of different words, scores for words, speech, repeated reviews, abandoned statements, and verbal rate. The original normative databases were created with samples from North Americans, but SALT developers say they are receptive to collaborations with other nations interested in developing normative data with SALT for their populations (OVERTON; WREN, 2014), (PEZOLD, IMGRUND; STORKEL, 2020).

7. Praat: Enables you to record or open audio files (aiff, wav or flac format) Free access. (http://www.fon.hum.uva.nl/praat/). It is a software for speech analysis and synthesis that was developed by linguists from the Institute of Phonetic Sciences, at the University of Amsterdam. It requires reading manuals and tutorials (available in English, Spanish, German, Swedish and Hungarian). Transcription of audio data can be performed with the Transcriber10 tool, which combines text and sound and creates annotation TAGS (CANAULT; LE NORMAND; THAI VAN, 2017), (GENEST; MASSON, 2017). Developed for the analysis of sound waves, with parameters such as frequency, length, intensity and formants. It is possible to apply the analyzes from files generated in other software (LENA, CLAN, Phon, ELAN), in order to extract, for example, information about the prosody of the speech performed or perceived by the child (ROSE; STOEL -GAMMON, 2015), (CANAULT; LE NORMAND; THAI VAN, 2017).

LENA, the software from which we obtained more records (10 articles) in this systematic review, is used to study early childhood language development in natural environments, with ecological validity. We only found one other software program (TimeStamper) with the same purpose. It was only used by the study group that developed it. LENA facilitates the collection and analysis of audio recordings of infants, providing automated measures for the speech heard and produced by the child. However, only the audio captured with the LENA recorder can be analyzed by the software itself. In a study by Canault et al., (2016) healthy French children from three to 48 months of age were recruited and divided into six age groups. These groups corresponded to one of the following stages of language development: vocalizations between 3 to 6 months; babbling between 6 to 12 months,

first words between 12 to 18 months, vocabulary explosion between 18 to 24 months, grammatical explosion between 24 to 36 months and grammar stabilization between 36 to 48 months. These age groups are exceedingly difficult to study with objective measures, which could explain the number of studies that have used this software.

The studies were classified according to the descriptions in the method section of each article: longitudinal (when they followed the development of language over a period), cross-sectional (studies that refer to language assessment at a given time), software assessment (studies that aimed to evaluate the software), and studies for software development (see Figure 3). All selected studies were observational or presented software reviews or development. Because they are not homogeneous, their statistical results could not be used to perform a meta-analysis.

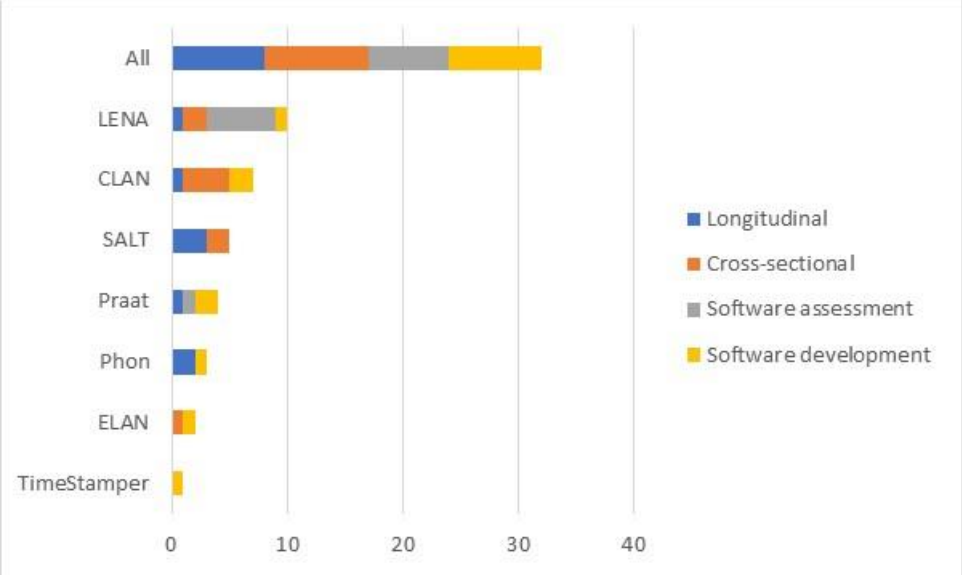

FIGURE 3 – Study design according to the method described in the articles included in this review.
Source: authors.

The software were used to analyze the speech of children with difficulties, as well as typical development. LENA was used in studies involving children with autism spectrum disorder (ASD), late language emergence (LLE), Down's syndrome, and chromosomal deletions. TimeStamper was used for children with cleft palates. SALT was used in research including children with hearing loss and ASD, whereas Phon was used for children with phonological disorders and hearing loss.

The reported methodology in the articles showed a variety of environments in which the recordings of spontaneous speech were collected. These settings included: homes, schools, outpatient clinics, and diverse environments (i.e., when the child used a tape recorder over a day or more).

The recording times for speech transcription with software similarly differed among the reported data. It is important to emphasize that some of the times presented were selected clippings from the transcriptions of recordings that may have been longer and perhaps carried out over

several days or for different periods in the child's development, as in longitudinal studies. The socioeconomic levels of the participants were assessed, but many studies did not report this data (15 of 24 publications). Those who included this information used different measures: while some evaluated class and income brackets, others assessed parental education and used questionnaires to gather information about house sizes and types of employment. This aspect of the research showed considerably heterogeneous data among the studies.

3. Discussion

The choice of software to transcribe the speech of young children depended on study objectives, the desired analysis, and the characteristics of the research population. Some software was used to focus on certain age groups, such as LENA and TimeStamper. These were chosen to examine prelinguistic vocalizations, babbling, and first words, in addition to capturing information about the linguistic diversity to which the child is exposed. Other software programs in this review also permitted a linguistic analysis of children who are already building words and longer utterances.

The software performed many types of analysis and assisted with investigations concerning phonetics and phonology, vocabulary diversity and sentence constructions (e.g. size and type of utterance and syntactic index), and even the prosodic characteristics of speech (such as intonation, babbling, and chaining) and gestures. For these reasons, some of the studies used more than one category of software in their analysis. For example, Burgess et al. (2013), used LENA to collect recordings and identify occurrences of adult words, and SALT to transcribe and analyse quantity, quality, and diversity of adult speech in a sample with autistic children. In another study, CLAN was used together with ELAN software. The first software was used for the analysis of phonological and morphosyntactic aspects, in addition to accounting for elements of speech production. The second software examined non-verbal communicative interactions and gestures (GOMES DE MELO BEZERRA et al., 2016). Most software offers the possibility to export data to other programs for additional analysis. The Praat software, designed to analyze the processing of acoustic signals, was used in studies with other software: LENA (BUSCH et al., 2018) (CANAULT; LE NORMAND; THAI VAN, 2017), CLAN (GENEST; MASSON, 2017) and Phon, (ROSE; STOELGAMMON, 2015).

Some types of software share similar functions but stipulate different conditions of access. Pezold et al. (2020) used two software for the same purpose in order to compare them. They chose CLAN and SALT to compute the MLU and NDW in the speech of two preschool children of different sexes and compared the results with those of other children from the same age group, available in databases open to the software. The results were similar regardless of the program used. The programs differed only when comparing the children's performance with a database of samples collected in different environments. This highlights the importance of standardized data for comparison purposes.

Some studies used databases provided by the software to prepare and develop analyses. For example, CLAN was developed using a standard transcription method called CHAT (Codes for the Human Analysis of Transcripts) to build a free international database that is called the CHILDES platform (Child Language Data Exchange System). CHAT has tools not only for the transcription of utterances (the so-called graphemic transcription), but also entire scenes of enunciation that may include additional notes referring to pragmatic elements of facial expressions, gestures, intonations, prosodic elements, morphosyntactic analysis, and any other necessary aspects (DEL RE et al., 2011). Like CLAN, Phon software has a multilingual corpus (PhonBank), designed to assist with the analysis of phonological and phonetic data transcribed with CHAT. There are full compatibility and interoperability with CLAN. PhonBank and TalkBank, in turn, serve as an expansion to the CHILDES platform and support research in language acquisition (BYUN; ROSE, 2016).

Another point to consider is the software language support. According to the information presented in Table 2, studies carried out with preschool children have not been done in many languages, although most of these software programs do not limit their use only to the language in which they were developed. Since languages around the world are morphologically distinct, a grammatical analysis requires an understanding of a specific corpus of the language spoken in the place of study. In Brazil, a research group from the Laboratory of Emerging Linguistic Productivity Laboratory (LAPLE) at the Federal University of Santa Catarina has been creating a program for the morphological analysis of Brazilian Portuguese with CLAN software, using the CHILDES database. However, this type of analysis is still not available in Portuguese. A preliminary analysis of data extracted from the "pau003.cha" corpus - which consists of 10,688 utterances in Portuguese recorded by adults and children - is available for use (SCLIAR-CABRAL; VASILÉVSKI, 2011).

Each language has specific phonetic, prosodic, and acoustic characteristics. Therefore, this diversity makes it difficult to define parameters that may lead to a general, non-English count of linguistic units. Canault, Le Normand and Thai Van, (2017) found LENA validations for Chinese, Mandarin, and Spanish. Currently, work is underway for Korean, Arabic, and Vietnamese, as well as for bilingual and trilingual environments. However, these studies report different reliability rates and do not systematically follow the same methodology. In fact, these different validations are based on audio samples of different sizes, with varying samples and analysis parameters. AWC and NCV are the variables that show the best agreement and correlation rates. We mentioned previously that it is possible to check the agreement rates and correlation coefficients for studies in different languages.

The applicability of the software is focused on research area, not on clinical practice, according to the studies in this review. Skahan and Lof (2007 cited in SARAIVA *et al.*, 2017), found that only 8% of speech therapists used computerized phonological analysis procedures. According to these authors, this low percentage of use could be explained by a lack of access to computers with the necessary support, by the cost of some software programs, or by a lack of familiarity with the technology and/or knowledge regarding the availability of programs with computerized evaluation methods.

Overton and Wren (2014) conducted a pilot study with schoolchildren aged 5 to 12 years, using naturalistic language samples. The results showed that measurement in a natural environment is

often problematic. Additionally, these analyses are more time consuming and are currently considered unfeasible in the clinical setting. As a result, professionals generally resort to the use of formal assessments with standardized instruments. Despite this, the authors observed that sampling in a natural environment provided a more complete and realistic picture of a child's skills and suggested that software can facilitate this task. Moreover, they noted that reliability calculations between the transcripts of a speech therapist and those of an assistant were considered acceptable for most measurements of naturalistic language skills.

Busch et al. (2018) compared the analyses of manual transcribers using Praat and LENA software. The transcribers listened to each audio file for five minutes without knowledge of the LENA results. They used TextGrids in Praat to take notes, marking speech intervals for each of the speakers and electronic sounds separately (e.g. sounds from televisions or other electronic devices). The transcribers used different tags to mark uncertain words and lexical status (such as onomatopoeia, babbling, family language practices, word forms, neologisms, laughter, crying, pauses, and overlapping). A comparison of these analyses with those of the LENA algorithm - which measures pauses of five seconds or more to constitute the end of a conversation and pauses of 300 milliseconds to end a vocalization, and excludes vegetative sounds (e.g., screams and eructation) – showed that LENA can detect high discursive variability in early childhood, despite its lack of discernment of lexical and semantic language, which only human transcribers can properly analyse these domains (CANAULT; LE NORMAND; THAI VAN, 2017).

This review included articles that used free-access software, or others available through authorization or purchase. All of them have manuals and require some type of preparation or training before use. Thus, in addition to technological and computer support, working with these software programs involves learning and handling time and, at times, financial resources.

Certainly, many other studies are using the software mentioned in this systematic review, with other age groups or that did not use spontaneous speech. Other software that allows phonetic transcription may not have been included in this study because they do not comply with pre-defined criteria, such as studies that were in progress with only publication in annals, congresses, conferences, or those that did not comply with the applied descriptors or in the searched database. This systematic review made it possible to bring the main software used in research with pre-school children, which can help other researchers or professionals who look for evidence regarding phonetic transcription in this approach.

4. Conclusion

The results showed a limited number of software programs (7) being used to assist phonetic transcription, with differences in functionality reported in the 24 articles included in this review. The types of analysis carried out with the software included pre-linguistic and prosodic analysis, phonetic and phonological analysis, morphosyntactic, grammatical and semantic analysis, extra-linguistic, and gestural and contextual analysis. The software was used to analyze the collected speech of

children with speech and/or language learning difficulties, and also children with typical development, from different environments (homes, schools, and outpatient clinics).

Through this systematic review, it can be concluded that there are clear advantages to using software for phonetic transcription, sample storage, and child language analysis, especially concerning standardization and reliability for spontaneous speech samples. Most of the software was developed to meet the needs of the international scientific community.

Acknowledgements

We would like to thank Viviane Castanho, a member of the library staff at the School of Medicine at UFRGS. She readily assisted with the step-by-step development of this systematic literature review and the alignment between search strategies and study aims.

Additional information

Evaluation and authors' response

Evaluation: https://doi.org/10.25189/rabralin.v21i1.2057.R
Authors' response: https://doi.org/10.25189/rabralin.v21i1.2057.A

Conflict of interest

The authors declare that there are no conflicts of interest.

Research protocol and pre-registration

Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020191698

REFERÊNCIAS

ALVAREZ, A. et al. Automating live and batch subtitling of multimedia contents for several European languages. **Multimedia Tools and Applications**, 75, p. 10823-10853, 2015. DOI https://doi.org/10.1007/s11042-015-2794-z

BALCIUNIENE, I. Lithuanian narrative language at preschool age. **Eesti Rakenduslingvistika Uhingu Aastaraamat**, n. 8, p. 21–36, 2012. DOI: http://dx.doi.org/10.5128/ERYa8.02

BREDIN-OJA, S. L. et al. "Clinician vs. Machine: Estimating Vocalizations Rates in Young Children With Developmental disorders". **American Journal of Speech-Language Pathology,** v. 27, n.3, p. 1066–1072, August 2018. DOI https://doi.org/10.1044/2018_AJSLP-17-0016

BURGESS, S.; AUDET, L.; HARJUSOLA-WEBB, S. Quantitative and qualitative characteristics of the school and home language environments of preschool-aged children with ASD. **Journal of Communication Disorders**, v. 46, n. 5–6, 428–439, 2013. DOI https://doi.org/10.1016/j.jcomdis.2013.09.003

BUSCH, T. et al. Correlation and agreement between Language ENvironment Analysis (LENA™) and manual transcription for Dutch natural language recordings. **Behavior Research Methods**, v. 50, n.5, p. 1921–1932, 2018. DOI https://doi.org/10.3758/s13428-017-0960-0

BYUN, T. M.; ROSE, Y. Analyzing clinical phonological data using phon. **Seminars in Speech and Language**, v. 37, n. 2, p. 85–105, 2016. DOI https://doi.org/10.1055/s-0036-1580741

CANAULT, M. et al. Reliability of the Language ENvironment Analysis system (LENA™) in European French. **Behavior research methods**, v. 48, n.3, p. 1109–1124, 2016. DOI https://doi.org/10.3758/s13428-015-0634-8

CANAULT, M.; LE NORMAND M-T.; THAI VAN, H. LENA^{TM} (Language ENvironment Analysis System): An automated speech recognition system and child language environment. Enfance, v. 2, n. 2, p. 199-216, 2017. DOI https://doi.org/10.4074/S0013754517002038

CHAPARRO-MORENO, L. J.; REALI, F.; MALDONADO-CARREÑO, C. Wordless picture books boost preschoolers' language production during shared reading. Early Childhood Research Quarterly, v. 40, p. 52–62, 2017. DOI https://doi.org/10.1016/j.ecresq.2017.03.001

COLLETTA, J-M.; COLLETTA, C.; GUIDETTI, M. Age-related changes in co-speech gesture and narrative: Evidence from French children and adults. Speech Communication, v. 52, n. 6, p. 565-576, June 2010. DOI https://doi.org/10.1016/j.specom.2010.02.009

D'APICE, K.; LATHAM, R. M.; VON STUMM, S. A naturalistic home observational approach to children's language, cognition, and behavior. Developmental Psychology, v. 55, n. 7, p. 1414–1427, 2019. https://doi.org/10.1037/dev0000733

DEL RE, A.; HILARIO, R. N.; MOGNO, A. S. Programa CLAN da base CHILDES: normas de transcrição (CHAT) e comandos básicos In: G. FERREIRA-GONÇALVES, M. R., BRUM-DE-PAULA & M. KESKE-SOARES (Eds.), **Estudos em aquisição fonológica**. Pelotas: Editora e Gráfica Universitária PREC-UFPel, v. 4, p. 11-30, 2011.

FLIPSEN, P. JR.; KANGAS, K. Mean length of utterance (MLU) in children with cochlear implants. **Volta Review**, v. 114, n. 2, p. 135–55, 2014.

FREITAS, L. B. L.; SHELTON T. L. Atenção à Primeira Infância nos EUA e no Brasil. **Psicologia: Teoria e Pesquisa,** V. 21 n. 2, p. 197-205, Mai-Ago 2005. DOI https://doi.org/10.1590/S0102-37722005000200010

GENEST, C. S.; MASSON, C. The contribution of corpus linguistics to the study of clinical situations: using ecological resources. **Studii de lingvistica**, 7, 89–112, 2017.

GOMES DE MELO BEZERRA, J. T. et al. Softwares de transcrição como auxílio para as pesquisas com enfoque multimodal no processo de aquisição da linguagem. **Texto livre-linguagem e tecnologia**, v. 9, n. 1, p. 77–93, 2016. DOI https://doi.org/10.17851/1983-3652.9.1.77-93

GREENWOOD, C. R. et al. Assessing Children's Home Language Environments Using Automatic Speech Recognition Technology. **Communication disorders quarterly**, v. 32, n. 2, p. 83–92, 2011. DOI https://doi.org/10.1177/1525740110367826

IPA. International Phonetic Association. [cited 2021 Jun29]. Available from: https://www.internationalphoneticassociation.org/content/links.

JULIEN, C. et al. Morphosyntactic development and severe parental neglect in 4-year-old french-speaking children: ELLAN study. **Child Maltreatment**, v. 24, n. 3, p. 254–264, 2019. DOI https://doi.org/10.1177/1077559519829249

MOHER D. et al. The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. **PLoS Med** *v.* 6, *n.* 7, e1000097. DOI https://doi.org/10.1371/journal.pmed1000097

MOELLER, M. P. et al. Longitudinal development of phonology and morphology in children with late-identified mild-moderate sensorineural hearing loss. **Ear and hearing**, v. 3, n. 5, p. 625–635, 2010. DOI https://doi.org/10.1097/AUD.0b013e3181df5cc2

OLLER, D. K. et al. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. **Proceedings of the National Academy of Sciences**, v. 107, n. 30, p. 13354–13359, 2010. DOI https://doi.org/10.1073/pnas.1003882107

OVERTON, S.; WREN, Y. Outcome measurement using naturalistic language samples: A feasibility pilot study

using language transcription software and speech and language therapy assistants. **Child language teaching & therapy**, v. 30. n. 2 , p. 221–29, 2014. DOI: https://doi.org/10.1177/0265659013519251

PEZOLD, M. J.; IMGRUND, C. M.; STORKEL, H. L. Using Computer Programs for Language Sample Analysis. **Language, Speech, and Hearing Services in Schools**, v. 51, n.1, p. 103–114, 2020. DOI https://doi.org/10.1044/2019_LSHSS-18-0148

RÄSÄNEN, O. et. al. Automatic word count estimation from daylong child-centered recordings in various language environments using language-independent syllabification of speech. **Speech Communication**, n. 113, p. 63–80, 2019. DOI https://doi.org/10.1016/j.specom.2019.08.005

RATNER, N. B.; MACWHINNEY, B. Your Laptop to the Rescue: Using the Child Language Data Exchange System Archive and CLAN Utilities to Improve Child Language Sample Analysis. **Seminars in Speech and Language**, v. 37, n. 2, 74–84, 2016. DOI https://doi.org/10.1055/s-0036-1580742

ROSE, Y.; STOEL-GAMMON, C. Using PhonBank and Phon in studies of phonological development and disorders. **Clinical Linguistics & Phonetics,** v. 29, n. 8–10, 686–700, 2015. DOI https://doi.org/10.3109/02699206.2015.1041609

SABRI, M.; FABIANO-SMITH, L. Phonological Development in a Bilingual Arabic-English-Speaking Child With Bilateral Cochlear Implants: A Longitudinal Case Study. **American Journal Of Speech-Language Pathology**, v. 27, n.4, 1506–1522, 2018. DOI https://doi.org/10.1044/2018_AJSLP-17-0162

SARAIVA, D. et al. Paediatric Automatic Phonological Analysis Tools (APAT). **Logopedics, phoniatrics, vocology**, 42, 153–159, 2017. DOI https://doi.org/10.1080/14015439.2016.1237544

SCLIAR-CABRAL, L.; VASILÉVSKI, V. Descrição do português com auxílio de programa computacional de interface. In: Anais da II Jornada de Descrição do Português, Cuiaba, MT, Brasil, 2011, p.39-46.

VENTURINI, C. *et al.* Confiabilidade intra e interexaminadores de dois métodos de medida da amplitude ativa de dorsiflexão do tornozelo em indivíduos saudáveis. **Revista Brasileira de Fisioterapia**, v.10, n. 4, 407-411, 2006. DOI https://doi.org/10.1590/S1413-35552006000400008

WILLADSEN, E.; PERSSON, C.; APPELBE, D. A software program to assist coding of prelinguistic vocalizations in real time. **Clinical linguistics & phonetics**, v. 32, n.10, 972–978, 2018. https://doi.org/10.1080/02699206.2018.1463396

XU, D.; RICHARDS, J. A.; GILKERSON, J. Automated Analysis of Child Phonetic Production Using Naturalistic Recordings**. Journal Of Speech Language and Hearing Research**, v. 57, n.5, p. 1638–1650, 2014. DOI https://doi.org/10.1044/2014_JSLHR-S-13-0037

APPENDIX 1

Supplemental material to document about review process.

Steps on the methodology applied for the systematic review study entitled:

"Phonetic transcription of spontaneous children's speech with the aid of software: a systematic review"

1. Formulation of the research project, registration on the PROSPERO Platform, verification of originality of the research (https://www.crd.york.ac.uk/prospero/).

- PROSPERO International prospective register of systematic reviews: under CRD number 42020191698 on 07/26/2020. This certified that our research is unprecedented.
- Registered review title: Children's language development assessment with phonetic transcription via software: a systematic review

2. Review team members and their organizational affiliations:

- Débora Tomazi Moreira Caumo. Postgraduate Program in Child and Adolescent Health, Universidade Federal do Rio Grande do Sul/ UFRGS.
- Camila Tomazi Moreira Caumo. Integrated Residency Program in Health. Basic Health Program: Escola de Saúde Pública do Rio Grande do Sul, ESP/RS.
- Márcio Pezzini França. Professor, Department of Preventive and Social Dentistry, Universidade Federal do Rio Grande do Sul/ UFRGS.
- Clécio Homrich da Silva. Professor, Department of Preventive and Social Dentistry, Universidade Federal do Rio Grande do Sul/ UFRGS.

3. Review question: Is there software that assists in the task of phonetic transcription of children's speech?

4. State the sources: Previously published in journals, specialist magazines, or even indexed in Library Cochrane, PubMed e Web of Science.

Search dates: In the last 10 years (June 2010 to June 2020).

Period of search in the databases: 18/06/20 a 24/06/20.

Restrictions (e.g. language or publication date): There are no restrictions as to location and language.

5. Keywords: Speech, phonological, phonetic, transcription, computer, software, children, language development, child language assessment.

6. URL to search strategy:

- PubMed: (phonetic transcription*[tw] OR phonologic transcription*[tw] OR phonological transcription*[tw] OR (speech*[tw] AND transcri*[tw])) AND (Software[mh] OR software*[tw] OR computer*[tw] OR automated[tw] OR automation[tw])

- Cochrane Central Register of Controlled Trials: phonetic transcription* AND software*

- Web of Science: TS=("phonetic transcription*" OR "phonologic transcription*" OR "phonological transcription*" OR (speech* AND transcri*)) AND TS=(software* OR computer* OR automated OR automation)

7. Participants/population: Children of preschool age.

8. Intervention(s), exposure(s):

- Software that can assist and facilitate the task of phonetic transcription.
- Speech of children assessed through phonetic transcription and use of software to assist in this task.
- A phonetic transcription applied to spontaneous speech.

9. Types of study to be included: There are no restrictions on the types of study design eligible.

10. Main outcome(s): Find software that assists in performing phonetic speech transcription to assess children's language development.

11. Study selection:

- Two authorized independent reviewers select the articles according to the preestablished requirements.
- Studies involving phonetic transcription and the use of some software for this task was included. If there is NO certainty in the step of reading the title and the abstract, the article was included for next step.
- The Zotero reference management software was used by the reviewers, allowing Merging of selected articles at the end of independent selection step, which result in the inclusion of all selected by both reviewers without duplicates.

12. Data extraction: title, author, magazine, place, year, software used and in which task.

13. Risk of bias (quality) assessment: The selection criteria of the PRISMA protocol (Preferred Reporting Items for Systematic reviews and MetaAnalyzes) was used.

14. Strategy for data synthesis:

- The data was synthesized to allow the choice of software for performing phonetic transcription.
- Scientific articles from different types of designs included, since language studies and assessment tools can involve different methods.

- The data that influence the choice of software was synthesized: information about its usefulness, applicability and functionality.
- The PICOS qualitative data demonstration method applied:

Participants: Data on the children's age group, participants with or without pathologies, sample number.

Interventions: What software is used in phonetic transcription and its characteristics.

Comparisons: If comparative groups exist, they can be detected in the data display.

Study design and results: data related to the type of study carried out and the results related to the use of the software was synthesized.

15. Analysis of subgroups or subsets:

- There may be subgroups to separate between software that perform or recognize speech and perform the task of phonetic transcription, and between software that assist the transcriber but do not perform the phonetic transcription itself.
- The function for which the software was applied, and for which type of population (children with or without pathologies) was described.
- It was intended to summarize about the technological support: of the software: information about audio and video (WAV, MP3), necessary operating systems (Windows, Mac Os X, Linux).