Bringing the phonological Development of Italian Children aged 30-47 months into PhonBank

The present paper represents an important contribution towards the PhonBank Initiative (a computerized online database of recordings and transcripts documenting phonological development, e.g., Rose & MacWhinney 2014), by filling the gap on the availability of Phon coded data on the phonological development of Italian children aged 30-47 months. PhonBank and Phon are recent outgrowths of earlier work on the Child Language Data Exchange System (CHILDES) Project.

In 1984, the primary emphasis of CHILDES was on the shift from handwritten notes to computerized files that could be subjected to automatic searching for developmental patterns, but it was not until 1995 that transcripts began to be linked to digitized audio. Prior to this, in 1993, an initial attempt was done to extend the CHILDES system to work on phonology. However, the lack of a consistent method for encoding International Phonetic Alphabet (IPA) characters on the computer at that time was a stumbling block. In recent years, the advent of Unicode, XML, improved digital audio and affordable file storage have removed the remaining technical stumbling blocks leading to the establishment of PhonBank (a computerized online database of recordings and transcripts documenting phonological development). The overall goals of PhonBank are fully compatible with those of TalkBank (a system for sharing and studying a wide range of conversational and communicative interactions), as well as with new developments in corpus phonology. The basic idea is that, by constructing a large database of accurately transcribed data on phonological development, we can test alternative theories of phonology and phonological development. For example, we can examine the role of universals in phonological markedness versus influences from distributional language specific patterns in the input to the children. We can then use these baseline patterns from normal development to understand the range of variation in learning and ways in which children with language disabilities diverge from developmental norms. The basic tools of PhonBank can also be used to study phonological systems beyond early child language; these tools can be applied to the acquisition of second languages, the learning and mastery of dialects, and the study of phonological effects in aphasia and other communicative disorders.

Within PhonBank, phonological transcriptions and acoustic data can be elaborated thanks to Phon, an open-source software built to support corpus-based phonological analysis (Hedlund, Gregory & Rose, 2018. Phon 3.0.0 [Computer Software]. Retrieved from https://phon.ca, on 27/09/2018). The creation and development of Phon responds to the central need within PhonBank to have a mean to build corpora annotated by means of phonetic transcriptions and TextGrid annotations, and to analyse its data. PhonBank is unique because all of its corpora are fully transcribed in IPA, with most children documented between 10 and 36 months of age, which also includes relatively rare transcriptions of children's very early stages of production, including studies on babbling. Because of the young age of recorded infants who rarely produce spontaneous speech, PhonBank also includes some transcriptions of adults' inputs. In line with both CHILDES and the general TalkBank system (see below), another central feature of PhonBank lies in the idea of data sharing: linguists and researchers who utilize dataset or software developed within these systems are encouraged to contribute to its improvement and expansion. They can upload new data relevant to the PhonBank project using the guidelines provided from the Website http://talkbank.org/share/contrib.html. PhonBank is indeed part of a wider project called TalkBank, a large system that combines 17 different databases, each of which is specific of a particular branch (below called "banks") for the study of language. All these different data collections are central to current investigations in many different areas of language. PhonBank is the one specific for phonetic and phonological language development.

The general graphical interface of Phon is composed by a set of windows and each of them has a specific function. A typical operative display includes Media and Segmentation, Record Data, Syllabification, Alignment, and Waveform. Phon supports multimedia data linkage, unit segmentation, blind transcription, automatic syllabification of data (according to the rules of several languages, Italian included, thanks to ISTC-CNR, http://www2.pd.istc.cnr.it/FESTIVAL/home/default.htm), model (Target) and produced (Actual) phonological forms and their alignment. This alignment is useful to perform systematic, phone-by-phone

comparisons between model and produced forms. An essential function of Phon is the Query browser for data extraction and analysis. According to one's research question, different types of queries can be created to fulfill specific analyses, many of which can be used to identify and study speech errors, thanks to the system's capacity to match the aligned model/produced phones present in corresponding IPA Target and Actual tiers. In order to exemplify the potential of PHON, we will present some of the analyses carried out on the speech productions of a group of 30 typically developing children from the Veneto region in Italy (aged from 30 to 47 months, equally divided by gender) which we coded into Phon. The children were recruited at different times from 2011 to 2016 by graduating students of the logopedic course of the University of Padova, mainly in kindergarten or at the children's homes, in the Veneto region (I). The children's parents gave their informed consent, compiled the MacArthur CDI surveys for their children's productions (Caselli, Pasqualetti & Stefanini, 2007), from which we estimated the size of their vocabularies, and the parents also filled a questionnaire from which we inferred the children to be, among other things, of normal psycho-physical development and monolingual (Italian). The children were administered a new "Test Fonetico per la Prima Infanzia" (*Phonetic test for toddlers*" or TFPI, see XXX), which basically is a picture naming test. The children's speech productions were manually transcribed in IPA, by several skilled transcribers, which finally confronted with a second series of transcriptions made by the first author, until a final agreement was reached by consensus. Transcriptions were carried on with the help of PRAAT facilities, and textgrids were created with segmentations and labels which were later successfully exported into Phon, by means of a Praat script by XXX.

For the present proposal, we divided the children into 6 groups spanning three months each. Exploiting the functionalities available in Phon, we derived individual phonetic inventories as well as group inventories (based on age). A phone is considered to be acquired if it is realized at least 50% of the times it was proposed to the child within the target words (cutoff criterion). Beyond phonetic inventories, which is a product of an "independent" analysis of the children's corpus, since it does not assess the correctness of the productions, an "Accuracy" analysis was produced. This kind of analysis is qualified as "relational", because the phonetic segments in the words produced by children are systematically compared with the target words, and the target segments are classified as "accurate", "substituted", "deleted" (e.g. Rose & Stoel-Gammon 2015; McAllister Byun & Rose 2016). The results of the above analysis will be compared and discussed with reference to both a previous study of the same authors (XXX) on phonetic inventories of Italian children aged 18 to 27 months, and to existing studies on Italian children of comparable ages (Zanobini, Viterbori & Saraceno, 2012; Tresoldi et al., 2018). In addition to this, another corpus consisting in the speech productions of 10 Italian children recorded every three months from 18 to 48 months of age, will be uploaded to PhonBank in the course of 2019, thus providing the scientific community with additional reference data on Italian.

References

Caselli M.C., Pasqualetti P. e Stefanini S. (2007), Parole e frasi nel "Primo vocabolario del bambino", Milano, FrancoAngeli.

McAllister Byun, T., & Rose, Y. (2016). Analyzing Clinical Phonological Data Using Phon. *Seminars in Speech and Language*, *37*(2), 85–105.

Rose, Y., & MacWhinney, B. (2014). The PhonBank Project: Data and Software-Assisted Methods for the Study of Phonology and Phonological Development. In J. Durand, U. Gut, & G. Kristoffersen (Eds.), *The Oxford Handbook of Corpus Phonology* (pp. 380–401). Oxford: Oxford University Press.

Rose, Y., & Stoel-Gammon, C. (2015). Using Phonbank and Phon in Studies of Phonological Development and Disorders. *Clinical Linguistics & Phonetics*, *29*(8–10), 686–700. https://doi.org/10.3109/02699206.2015.1041609

Tresoldi, M., Barillari, M. R., Ambrogi, F., Sai, E., Barillari, U., Tozzi, E., Scarponi L. & Schindler, A. (2018). Normative and validation data of an articulation test for Italian-speaking children, *International Journal of Pediatric Otorhinolaryngology*, 110, 81-86.

Zanobini, M., Viterbori, P., & Saraceno, F. (2012). Phonology and language development in Italian children: An analysis of production and accuracy. *Journal of Speech, Language, and Hearing Research*, 55(1), 16-31.