The Open Handbook of Linguistic Data Management

Edited By: Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister

Citation:

The Open Handbook of Linguistic Data Management
Edited By: Andrea L. Berez-Kroeker, Bradley McDonnell, Eve Koller, Lauren B. Collister
DOI: 10.7551/mitpress/12200.001.0001
ISBN (electronic): 9780262366076
Publisher: The MIT Press
Published: 2022
32 Managing Phonological Development Data within PhonBank: The Chisasibi Child Language Acquisition Study

Yvan Rose and Julie Brittain

1 Introduction

Since 2006, the PhonBank database project (https://phonbank.talkbank.org) has provided computer-assisted methods for the management and analysis of child language phonological data. These methods are assembled within Phon (https://www.phon.ca), an open-source software program that supports all of the data annotation standards required for the building and analysis of PhonBank corpora. PhonBank is a component of the larger TalkBank database system (https://talkbank.org), alongside CHILDES (the long-standing Child Language Data Exchange System; MacWhinney & Snow 1985) and other specialized databases such as FluencyBank and AphasiaBank. Each of these projects, as well as their continuous funding, has historically relied on public granting agencies as well as on significant levels of commitment within the research community. Just as scholars must use computer programs such as Phon to build and analyze their corpus data, the very existence of these programs and databases such as those of TalkBank is fully dependent on data sharing, which effectively transforms each corpus-building endeavor into a long-term investment for research and associated activities (e.g., educational, clinical), as we will exemplify.

Within the TalkBank family, PhonBank is unique in its organization around the Phon software program; all of the other databases center on the CLAN program https://(dali.talkbank.org/clan). While Phon and CLAN offer similar basic functionality (e.g., time alignment; standardized annotation system; query functions), Phon differs from CLAN in that it offers specialized functions for the study of phonetics and phonology (Rose & MacWhinney 2014). While these functions were originally designed and implemented to study phonological development, they can also be used to study virtually all topics related to phonology and acoustic phonetics (e.g., speech disorders, dialectal variation, sociophonetics, or field studies, among many others). In a nutshell, Phon is a database software program that integrates several tools for textual, phonological, and acoustic analysis. Phon also incorporates dedicated functions for the study of language acquisition and language disorders. For analyses based on textual forms, Phon can be used to encode text and incorporate lexical or morphosyntactic annotations, each of which is stored in dedicated data fields. These annotations can then be used as search criteria within powerful yet easy-to-use query functions to extract precise information relevant to the analysis at hand. Concerning phonetics and phonology, Phon automatically builds the phonological data structure from phonetically transcribed words. The transcriptions are internally analyzed by the program’s specialized algorithms to identify all phones, phonological features, and positions (within the syllable, word, relative to word stress, and so on). Phon also fully integrates with Praat software for acoustic analysis. Already existing Praat (TextGrid) data can be imported into Phon, and new TextGrids can be generated directly from within Phon. Given these combined functions, and the reduced need to manually enter phonological annotations into the corpus, we can perform an unlimited number of phonological and phonetic (acoustic) analyses, the results of which can be interpreted in light of all other information contained within the Phon database (e.g., speaker information, phonological context).

In this chapter, we describe the general methods of corpus building and analysis available within Phon, contextualizing our account by taking the Chisasibi Child Language Acquisition Study (henceforth, CCLAS) as a primary example. CCLAS is particularly relevant to the present Handbook in that it combines descriptions of the phonetic, phonological, and morphosyntactic units.
of Cree, a polysynthetic (Algonquian) language, as well as children’s recorded behaviors during their acquisition of these linguistic units.

2 The Chisasibi Child Language Acquisition Study

The primary aim of CCLAS, a community-initiated project that began in 2004, is to document the acquisition of Cree as a first language. The project began following some years of discussion between linguists at Memorial University and members of the Cree community in Chisasibi, with whom our department had a long-standing research relationship. The Cree School Board was particularly invested in the dialogue, having a nuanced understanding of the potential applications of the research, and they made the formal invitation to initiate CCLAS; they have been our community partner for the past fifteen years. Detailing the project’s practical applications is beyond the scope of this chapter; suffice to say, offering support to the community’s speech-language pathologists is a priority. Within the community there was a concern that more children than the expected average were being diagnosed with speech-language problems, an overdiagnosis, some felt, resulting from screening in a second language (English or French) due to the lack of Cree language tools; clearly, adequate first-language (Cree) screening tools were required to begin to address the issue. We say more about this herein.

CCLAS falls within the general tradition of naturalistic, longitudinal studies of language acquisition, which consists of samples of child language productions obtained at regular intervals through the recording of linguistic interactions within the child’s regular environment. The long-term objectives of CCLAS are to contribute to current debates in contemporary linguistic theory, especially within the subdiscipline of first language acquisition; to contribute to the descriptive literature for Cree, an underscribed and potentially endangered language (Wurm 1998; Brittain & MacKenzie 2016); and to assist in maintaining the vitality of the Cree language by sharing research findings with educators and speech-language clinicians who work in the Cree-speaking environment. Research published to date has been primarily based on our data sets from two children, code-named “Ani” (child A1) and “Billy” (child B3). Further information can be obtained from the project’s website (https://www.mun.ca/cclas).

CCLAS is unique in that it is, to our knowledge, the first and only research project on the acquisition of an Algonquian language. CCLAS is also the only project on the acquisition of any Indigenous language to contribute data to a publicly accessible database. A subset of Ani’s data was simultaneously contributed to the CHILDES and PhonBank databases in 2013. We are currently working toward expanding this data set with the publication of the other children’s data.

Cree is typically spoken in multilingual contexts; in Chisasibi, for example, most people are fluent in English and in French to a lesser extent, and both languages are used in domains such as education. We have thus utilized a set of annotations to identify code switches, which help to contextualize aspects of both phonological and morphosyntactic development, for example, concerning the development of phones and phone combinations that are unique to either the Cree or the English phonological systems (Bryant 2013). Finally, and perhaps more relevant to the purpose of the current chapter, the CCLAS database combines phonetic, phonological, and morphosyntactic annotations structured within Phon in ways that facilitate the parallel study of different aspects of the Cree language. As a result, phonological analyses can be informed by morphosyntactic observations, or vice versa, using a unified set of data transcriptions and annotations.

In the following sections, we describe how we proceeded with the most central aspects of this research and discuss how we responded to the many challenges it raised along the way. It is worth noting that CCLAS and the larger context within which it exists, namely Phon and PhonBank, have all emerged in virtual synchrony and have been tightly intertwined from the outset because all three projects constitute the ongoing work of researchers and computer scientists at Memorial University. In this regard, CCLAS has served as both a source of inspiration and testing ground for many of the annotation and query methods currently available within Phon.

3 Managing phonological data in a first language acquisition study: The CCLAS experience

As we alluded to, our building of the CCLAS database came with a unique set of challenges, which we tried to tackle in a pragmatic way, following the “agile” approach to corpus building highlighted in Voormann and Gut
(2008). Under this approach, corpus building consists of a series of sweeping data processing steps that apply over the full corpus in a way that ensures that methodological adjustments, unavoidable in virtually all corpus-building endeavors, are applied to the entire data set in a uniform fashion to ensure corpus and analytic consistency. We describe the most essential steps we took toward the building of our Cree data set, which together have enabled us to obtain interrelated observations (phonetic, phonological, morphosyntactic) about the children’s language performance throughout extensive developmental periods, observations that are supplemented with additional information (e.g., situational), helping us interpret these data within their proper context. We then describe some of the ways in which we have been disseminating our research outcomes within the Cree community and for scholarly research.

3.1 General context
Since its inception, CCLAS has been operating in two main locales, Chisasibi, Québec, the community in which our research is situated, and Memorial University in St. John’s, Newfoundland, where most of the work on data transcription and analysis has taken place. Fieldwork sessions with the project-affiliated Cree language consultants take place at both of these research sites. Another defining characteristic of this project, and principal driver of the methods we have developed at different stages in our research, is the fact that while Algonquianists specializing in Cree and related languages have been involved in the project since its inception, none of the team members based in St. John’s is a native speaker of Cree. In the following sections, we describe how we have adapted to these challenges in building the CCLAS database and in analyzing it. As we will see, conducting first language acquisition work on an underdocumented language imposes the double task of performing fieldwork to describe the grammatical and phonetic aspects of the Cree (adult) system being acquired by the children and then analyzing the children’s patterns of language development in light of the descriptions obtained from fieldwork. This is the only way to ensure that we interpret the children’s patterns of language development in their proper contexts.

3.2 Data collection
The CCLAS database comprises ninety-seven audio-video digital recordings made over a thirty-month period (November 2004 to April 2007), which cover the language-learning journey of six Cree-speaking children. These child participants fall into two age cohorts: the children from cohort A (children A1, A2, A3) were approximately 1;8 when recording began, while the children of cohort B (children B1, B2, B3) were approximately 3;6. Three of the six participants yielded corpora of a substantial nature, that of child A1 Ani (thirty-seven recordings), B1 “Daisy” (thirty-two recordings), and B3 Billy (nineteen recordings), amounting to approximately fifty-six hours of recording. The children were filmed by a Cree-speaking resident of Chisasibi who also served as CCLAS’s on-site project manager. During the recording sessions she was also the children’s caregiver, engaging with them in activities that elicit language—playing with toys, talking about recent activities, and so on; she is identified as “the adult” interlocutor in all the recordings. While recordings took place primarily in Cree, a minority of interactions also occurred in English, in contexts when the children preferred to use English words or phrases.

3.3 Data processing
The processing of these original recordings then followed a series of steps that took place at either of our sites, as will be described, and involved a combination of linguists and Cree language consultants. We summarize these steps herein.

Data segmentation consists of the identification of the time intervals on these recordings and their associated utterances. Each session recording received from Chisasibi is first identified with the date of the recording as well as the names of the participants. Using the segmentation function in Phon, every utterance recorded is then time-stamped and associated with its corresponding speaker (e.g., the child participant or the adult interlocutor), a task that can only be performed by someone who understands what the speakers are saying, necessarily, that is, a Cree speaker. To achieve this step in the processing sequence, the Phon data files are transferred to the Cree language consultant in Chisasibi, who works within Phon to do the segmentation.

After the segmentation step is completed, the Cree consultant provides an orthographic transcription (using roman script) and an English translation for each utterance (record). Translation is a key requirement bearing in mind the fact that none of the researchers are fluent
in the language. The updated files are then sent back to the researchers at Memorial University.

We then engage in the phonetic transcription of each child utterance. This step is rather difficult, primarily due to the fact that we do not have access to native Cree speakers trained for the task. To further alleviate some of the shortcomings related to the transcription of a language one does not speak (e.g., potential misperception of phonetic categories), we provide dedicated training on Cree transcription to undergraduate and graduate research assistants in linguistics who are native speakers of English. In addition, we rely on the double-blind transcription protocol, followed by consensus-based validation of the transcripts, as follows: Two independent transcribers first perform their phonetic transcriptions within dedicated interfaces, each without access to the other’s work. After they have completed their individual transcriptions, the same transcribers then work together to compare each of their respective transcriptions, select the one deemed the most valid, and, whenever needed, improve the selected transcription with additional details noted during the comparison. In cases where the transcribers cannot reach a consensus, even with the help of a third person or through other methods such as spectrographic inspection of the signal, the records are excluded from further research.

This brings us to the stage of data processing. Either in Chisasibi or in St. John’s, the team members work together to better understand the child language captured in the videos and to audio-record the “target” (adult) forms corresponding to each child utterance. This target form provides a baseline to assess the child’s productive abilities. The different kinds of information gathered in these work sessions, which generally run for about a week, are entered into the appropriate tiers in Phon. With regard to better understanding the language, the principal goals here are to clarify linguistically relevant context, to identify “incorrect” child productions, and to discuss grammatical constructions unfamiliar to the linguists. We now consider each of these in a little more detail.

For the non-native–Cree-speaking members of the team, understanding the context of the language, particularly the child utterances, is challenging because Cree has very liberal argument omission patterns so that knowing who is doing what to whom, so to speak, is dependent on being able to track complex functional forms, bound and free, across what can be relatively large spans of discourse. Viewing the recordings as a team, contextual information is clarified and observations are recorded in Phon. The Cree consultant also identifies cases where the form a child produces is not “on target,” differing from the adult form in some manner (phonology, morphology, syntax, lexical choice, and so on). Target forms are audio-recorded for all the child utterances and these are transcribed phonetically using the International Phonetic Alphabet (IPA) and entered into Phon by research assistants. For each child utterance (which corresponds to a data record within Phon), we have an IPA representation of the target and actual forms, allowing us to identify cases where the child falls short in her production. While this comparison can be done within Phon without the help of a Cree speaker, reviewing each utterance with the consultant provides invaluable information regarding, in particular, types of grammatical error or potential factors that might yield the errors observed; in some more extreme cases, an “error” may only be apparent but not actual in that it simply reflects the types of phonetic patterns that occur in familiar or frequently used expressions, aspects of which fly in the face of prescriptive language descriptions (e.g., the pronunciation of potato as [pəteɾ] instead of [pəter] in English). Finally, because Cree in general, and Northern East Cree in particular, is underdescribed, as we discuss both child and adult language, we frequently identify areas of the grammar where further fieldwork is required. The more theoretically oriented work undertaken on the project is dependent on the descriptive; for example, the adult stress system Northern East Cree was detailed within the context of the project (Dyck, Brittain, & MacKenzie 2006), work that facilitated a case study of the acquisition of this area of the grammar (Swain 2009). To disseminate this descriptive work more widely, our findings are also published on https://www.eastcree.org, a website oriented toward a Cree-speaking audience.5

The orthographic transcriptions provide the basis for our morphological breakdown of the utterances, which is performed by members of our research team with the relevant expertise. We input the morphological and related semantic information into dedicated fields within the Phon database, with the aid of a parser built specifically for this work and integrated into Phon. The screen shot in figure 32.1 illustrates the outcome of this work.
Managing Phonological Development Data within PhonBank

3.4 Data analysis

As discussed extensively in Rose and Inkelas (2011), beyond the work involved in the building of a longitudinal corpus of child language development, data interpretation arguably represents one of the most difficult challenges for any acquisition study. This is especially true in the context of CCLAS, given the frequent need to engage in fieldwork in order to set a basis to interpret the child data. Following the needs of the research at hand (e.g., on Cree speech phonetics, phonological or morphological development), we work in concert with graduate students and Cree language consultants to obtain the required data descriptions, which we maximally incorporate within our database, where they become not only useful in the current context but also available for future research. Also maintaining a pragmatic approach to database building, we typically expand our annotations on a need base and restrict annotation work only to the set of data records relevant to the study at hand. While this approach gets in the way of exhaustiveness for certain types of annotations, it also makes possible analyses that would otherwise be too time-consuming; this approach also optimizes research output relative to database-building time.

After all the necessary annotations are entered into our database, we proceed with data analysis. Phon supports different methods for database mining. To make powerful queries readily available to linguists, which would otherwise call for expertise in programming or scripting, we designed a series of query forms such as the one in figure 32.2, which provide intuitive guidance but leave much of the technical (programmatic) details out of the interface.

For example, one can look for particular morpheme types (e.g., the preverb, represented by the “pvb” text expression in figure 32.2) and display all the corresponding morphological or phonetic data within and across session transcripts, either for all participants or for a subset of the participants.

Similarly, we can look for a given phone class (e.g., obstruent stops) and study its development over time, within and across positions within the syllable or word; each time developmental or otherwise variable patterns are detected in the data, further queries can be formulated to identify the origins of these patterns. The current version of Phon also incorporates general measures such as the PMLU (Ingram 2002) or the percentage...
of consonants correct (Shriberg et al. 1997) as well as “detector” algorithms for phonological patterns such as fricative stopping or consonant harmony. These measures, which are commonly used in research on phonological development and speech disorders, can also be applied in the context of research on second-/foreign-language acquisition or dialectal variation or, in the case of the PMLU, to obtain broad measures of phonological complexity and related productivity.

Finally, to engage in acoustic analysis (beyond spectrographic visualization, readily available in Phon), we must perform a few additional preparatory steps. The first of these consists of obtaining TextGrids, which can be generated from within Phon to immediately incorporate tier data such as orthographic and phonetic transcriptions of the forms produced by the speaker. These transcriptions, in turn, serve as the basis for TextGrid alignment, where word, syllable, phones, or other measurable phenomena such as pauses serve as useful labels for the time intervals of the audio recordings that they represent. Although we are planning on integrating technology to automate the TextGrid alignment process, the current version of Phon only supports manual alignment of the TextGrid data. After TextGrid alignment is complete, we then identify the phones or phonological contexts to be measured through regular (orthographic or phonetic) data queries. After completion of a query, we can run one or more acoustic measurements on the data returned by the query. For example, one can look for broad measures such as pitch measurements on word forms; one can also employ narrow acoustic measures targeting specific phones such as high vowels produced in stressed versus unstressed syllables, and compare vocalic production across these two prosodic contexts based on a combination of formant, duration, intensity, or pitch data. (Other analyses include voice onset time and spectral moments, which are useful to study the production of obstruent stops and fricatives, respectively.)

The interface to select the desired analyses is illustrated in figure 32.3. For each acoustic analysis, we can set specific parameters, following all the methods readily available in Praat, which we incorporated within Phon in close collaboration with the developers of Praat. Because of the integration between Phon database and Praat acoustic measurement functions, acoustic data can be easily combined with other annotations within the database. For example, still keeping the general mixed-language context of the Cree community in

---

Figure 32.2
Example of a query form.
Chisasibi, we can compare phonetic units of production found in native Cree words with those observed in the English loanwords documented within our database.

Every report from textual, phonological, or acoustic analysis generated by Phon is displayed within a dedicated Phon window. From this window, the reports can be saved in a number of formats, again depending on the current need. Print-ready reports can be saved in HTML format. More typically in the context of scholarly research, Phon reports can be saved for postprocessing as Excel spreadsheets and workbooks, or exported as CSV files, for use in Open/LibreOffice or within most statistical software packages (e.g., SPSS, R).

3.5 Results dissemination and anonymized data sharing

As important as the dissemination of scholarly findings is our commitment to share our research outcomes with the Cree community; to this end, team members make regular presentations to educators and speech-language pathologists working within the community (e.g., Brittain et al. 2005; Brittain & MacKenzie 2010a, b) and to the community at large (e.g., Brittain, Johansson, & Rose 2011; Johansson 2012b), including being guests on the Cree-medium community radio. We are also committed to finding practical applications that support the language, and, to date, these contributions have been in two principal areas: supporting the work of speech-language clinicians (Brittain 2014) and contributing to the documentation and description of the language. As we have said, the active CCLAS database consists of three corpora, for which reason our output takes the form of case studies of various areas of the grammar. Although the database lacks the statistical power required to create screening tools for speech-language pathologists, which typically rely on normative data obtained...
from larger-scale sampling, we have proceeded on the assumption that the in-depth analyses we obtain from longitudinal case studies can provide useful information toward the building of these tools.

We maintain documentation of these outcomes on the CCLAS website, which also includes academic works on the acquisition of morphosyntactic (Terry 2010; Rose & Brittain 2011; Johansson 2012a) and phonological categories (Swain 2009; Rose et al. 2010; Bryant 2013), phonetic studies of the Cree language independent of its acquisition (Dyck, Brittain, & MacKenzie 2006; O’Neill 2014) as well as systematic comparisons between the children’s productions of native Cree words and their usage of English loanwords (Bryant 2013; Pile 2018). Beyond their contributions to theories of linguistics and language acquisition, these studies offer many resources toward educational or clinical initiatives as well as additional information to the community about multilingualism or more practical considerations such as that of raising a child within a mixed-language linguistic community.

Returning to the larger context of CHILDES and PhonBank, we also began our work toward the sharing of our corpus data with our research community. Similar to our corpus-building initiative, our initial corpus publication posed a number of questions about how to maximize data output while at the same time keeping within ethical boundaries in the areas of data anonymity and cultural propriety. Key to this was the removal of any contextual identifier so that members of the Cree community unaware of what particular children participated in our study would have no way to trace the data back to any individual child or related family. As video can never offer full anonymity, an early decision was to limit media publication to the audio tracks of the videos we recorded with the children. Beyond this obvious decision, a trickier challenge was that names of people and places are often mentioned as part of verbal interactions with the children, in particular these individual children’s names and that of their relatives. This is one of the drawbacks of the naturalistic approach; avoiding this type of issue is much easier through guided data elicitation (e.g., picture- or word-naming tasks). To remove this information, we first perused through our data transcripts to identify all of the names and dates present in data records. Using the time stamps of these records as a starting point, we then identified the exact time intervals containing these identifiers and replaced the sound information with “pink” noise, which can be described as a non-aggressive, low-intensity version of white noise, making the listening of these intervals, which are often found within larger utterances, more comfortable to the listener than the white noise itself. After each media file was fully deidentified using this method, we then proceeded with the replacement of the names within the corresponding transcript by meaningless strings such as “nnn,” to which we can still attach morphological or other types of labels as a unit of morphological analysis, also making sure to remove the IPA Target and Actual forms corresponding to the original name. Using a similar method, we were also able to remove names of places and dates, including the participants’ birth dates and the recording dates present in the original transcript metadata. We only left child’s age (at recording time), a time measure in the absence of corresponding calendar dates, as a means to map developmental data over time. Finally, in the face of potential concerns about one’s ability to identify speakers based on the characteristics of their own voice, we add the observation that given the changes that toddlers’ voices undergo during childhood, it is unlikely that someone other than, possibly, direct family members (who are already aware of the study), can identify the child speakers, especially given the relatively long period between the time of the original recording and the time when the (anonymized) data are uploaded to the online database (over a decade for the first subset of our corpus).

Resulting from this work is a data set that remains fully analyzable from both linguistic and developmental perspectives, but which is otherwise impossible to trace back to the original child speaker. This level of deidentification, and the relatively easy method employed to obtain it, makes the documents conform to the letter of our informed consent; it is also perfectly sufficient to meet the requirements of the EU General Data Protection Regulation (https://gdpr.eu), arguably the most restrictive international policy statement to date in the area of confidentiality for electronic documentation worldwide.

We then submitted our corpus to CHILDES and PhonBank through the TalkBank contribution system (https://talkbank.org/share/contrib.html), where it was curated and then published in both Phon and CHAT (Codes for the Human Analysis of Transcripts) formats, the latter being the format used by CLAN. The curation steps were minimal in this case, as the original corpus
was already within the Phon format, which is fully compliant with the general standards of TalkBank. In addition to the media and transcript files, corpus submission only requires a summary description of the corpus as well as a short list of references. According to the general sharing rules of TalkBank, any individual who benefits from the published data has an obligation to cite one or more of these references to the original authors’ works as part of any new study that builds on shared data. This view of data sharing and referencing is essentially an honor system that has supplied researchers and students with invaluable resources and methods since its inception in the early 1980s (MacWhinney 2000; MacWhinney et al. 2012; Rose & MacWhinney 2014). It also provides corpus contributors with an additional window to publicize works based on their original data set.

4 Conclusion

We have provided an overview of the building, analysis, and dissemination of corpus data on the first language acquisition of an Indigenous language. This work, still ongoing, has raised a number of difficulties, primarily due to its setting across multiple sites and the absence of a native speaker of Cree among our team in St. John’s. We have streamlined a methodology whereby data processing is undertaken in these different locales, by the non-Cree speakers in St. John’s, by the Cree speakers in Chisasibi, and, crucially, by these two groups meeting on a regular basis and working together to pool their expertise. We have also offered an overview of the types of analyses we can run within Phon, as well as some of the most important steps toward anonymous corpus data publication on PhonBank (or other similar databases within TalkBank).

In conclusion, we would like to highlight that very little of the work we described would have been possible without all the types of sharing we benefited from at every step along the way. As several people have shared their time, knowledge and expertise since the beginning of CCLAS, the whole TalkBank architecture for data sharing also rests on the premises that sharing effectively turns the large amounts of resources spent on corpus development and analysis into long-term investments into research and society more generally. In this context, it is our hope that our publication of Cree acquisition data will encourage similar initiatives with additional Indigenous languages, ideally within the context of TalkBank. This would yield many opportunities for cross-language, cross-corpus comparisons, thereby opening additional windows of understanding into each language involved and their acquisition.

Notes

1. The early development of Phon was funded by grants from the Social Sciences and Humanities Research Council of Canada, the Canada Fund for Innovation, as well as by a Petro-Canada Award for Young Innovators. Since 2006, the development of Phon and PhonBank has been funded primarily through grants from the US National Institutes of Health (R01 HD051698, R01 HD051698-06A1, and R01 HD051698–11).
2. The CCLAS database is Northern East Cree (ISO 639-3: crl).
3. CCLAS has been funded primarily by the Social Sciences and Humanities Research Council of Canada (Standard Research Grants 410-2004-1836 and 410-2008-0378 and Insight Grant 435-2013-1297), internal funding from Memorial University, and through financial and in-kind support from institutions within the Cree community, namely the Cree Nation of Chisasibi, the Cree School Board of Québec, and the Anjوابa Childcare Centre in Chisasibi.
4. In 2006, child A3 was diagnosed with atypical language development; for the present time, we exclude this child’s nine recordings. Children A2 and B2 were withdrawn from the study after only four recordings were made for each.
6. Users with the relevant level of expertise can also align the Phon-generated TextGrids using forced-alignment systems such as those built within LaBB-CAT, the Montreal Forced Aligner, or MAUS. More information about these and other forced aligner technologies can be found at https://github.com/pettarin/forced-alignment-tools.
7. We owe special thanks to Dr. Paul Boersma for his collaboration on this aspect of Phon development.

References

Brittain, Julie, and Marguerite MacKenzie. 2010a. The Chisasibi Child Language Acquisition Study: Bridging the gap—How to
make our data serve a practical use. Presentation to clinicians, Montreal Children’s Hospital, May 27.


Brittain, Julie, Marguerite MacKenzie, Carrie Dyck, and Yvan Rose. 2005. First language acquisition and the Chisasibi Child Language Acquisition Study (CCLAS). Presentation to the Board of Directors of Anjabowa Childcare/Headstart Centre, Chisasibi, April 25.


Terry, Kevin. 2010. The emergence of intransitive inflection in Northern East Cree: A case study. MA thesis, Memorial University of Newfoundland.
