Processing Child Language Samples

Clifton Pye pyersqr at ku dot edu

February 2021

This document describes the steps for producing a basic analysis of a child's language sample. The procedures start with a transcription in ELAN of a recording of a child interacting with other speakers. The analysis procedures export the child's utterances from ELAN to Excel via Unicode text files. The different steps in the process assume that the input files have a standard format. The document describes the use of Perl programs that create a lexical concordance (concord.pl) and a vowel concordance (vowel.pl).

The ELAN transcription

The ELAN transcription tool enables transcribers to link their transcription to audio and video files. These links make it easy to replay a selected portion of the recording in order to check the transcription. ELAN provides horizontal tiers for each speaker. These tiers make it easy to show the overlaps between the utterances of different speakers. The ELAN program can be downloaded from the Max Plank Institute for Psycholinguistics at https://archive.mpi.nl/tla/elan/download. Figure 1 shows a screenshot for a typical ELAN transcription.

Figure 1. Screenshot of ELAN transcription

				ELAN 5.1 - PJL27	1022016 A.maf				- 0	
e Edit Annotation Tie	Type Search Yew	Options Window	n Rindo							
arid Text Sublities	Lexicon Commente	Recognizers	Metadata Controls							
M 58 892		12200	00 34 58 892 - 00 35 00 151							-
4 14 14 14 14 -H	. INCOMENTATION				te 🗌 Loop Mode 🌒					
and the first of	e ter ter ter ter ter	THE DECTS								
127022016 A - 3 30.00	00 23 32 000	00 23 34 000	00 23 38 000	0 23 38 000 00 3	13 40 000 00 23 42 000	00 23 44 000 0	0 23 46 000	00 23 48.0	00 00 23 50 000	
-++			-++	- + -			• •	***	** * * ***	
		1.0000200		1000				100202		
						0.000	•	***		
										-
3.30.00		00:23:34.000					0.23.46.000	00-22-48.0		
Detault										
JL	kiua kut tsanē	masat	kua tsané kuatsa		koasnē niš7 s masat		ajaaa	aui	-	
190	kiua kut tsan78	masat	kua tsa7nê chin k		kua sa7ně ň7iā7 5 masat		2(222	x		
Equi-JL										
Trad-JL	adri oga kez kju	o venato	aqui sa vino aqui		aca se viene este venadot,		01 0 0 0 0			
Mama de JL	4			y 6757 5 gu		S matát no	guing	-	entonces ira kyi ntju7	
[AN]	1			y este Sgrot		venadote n	tore		enfonces mira armadili	
Trad-Marnä de JL	1									
Félar						Innn	-			
Trad-F dita						100				
Det										
Niño 1										
Equ-Milo 1										
					1				11-0-	
Transation										

The ELAN transcription shows the standard format for transcriptions of child language samples. The transcription provides one tier for the child's utterance (JL), a following tier for the adult equivalent of the child's utterance, and a third tier for a translation into another language. The analysis procedures assume that children's utterances have these three tiers and no others. The transcription shows that the child's utterances are only divided into words. The following two tiers show the typical transcription format for an adult speaker. The first tier shows the adult's utterance, and the second tier shows the translation. Adult utterances, like the child's are only divided into words. The analysis programs assume that the transcriptions have translation tiers, but do not require them. A blank translation line can be added at a later stage if needed. The transcription format follows the logic of the minimal transcription method described in Pye (2020).

Transferring ELAN transcriptions to Excel

The analysis programs require text files in the UTF-8 format. The first step in the analysis

X.		Toolbax File (UTF-8)	ELAN 5.1 - PJL27022	016 Aleaf			- 0
[ile [dit Arnotation	Tier Type Se	FLEX FRO	Пар				
Bew.,	CHA	CHAT File	Metadata Controls				
Open	2004	Isl-detenited Text.					
Open Becent File		Tiger.xml	0.34.58.892 - 06.35.00.151 1259				
Çkəse	Carra	Interfacear Export as tab delive	100 100 + - + + T - Selection Mode [Loop Mode			
Save	DH	HIML.					
Save As	CHILDRAS	Traditional Transcript Text					
Save as Template Save Selection as .eat.		Eraat TextGrid List of Words	00 23 36 000 00 23 38 000 00 23 40.				9 ka 000 00 23 80.000
Save a Copy as EAF 2.7		SMIL P					
Herge Transcriptions Automatic (Jackup		GuickTime text Sybittles Text Tiors for Recognizer					
Page Setgp Print Preylew	Cal-Sid-P Cal-NAP	Media Clip using M2 edit Media Clip using Script		- • · · • • • •			
Print	Calif	Image from ELAN Window	00 23 36 000 00 23 36 000 00 23 43	000 00.23.42.000 0	0 23 44 000 00	23 44.000 00 2	3 48 000 00 23 60 000
Multiple File Processing		Elimetrip Image					
Export As		Shoebox File	liua trani kuatra l	Ikoasné niš7 s masat		sisaa aui	- M1 - M1
Export Multiple Files As	, ,		Plua Darre Roada	KUWDIN MAY 9 MASA		11464 000	-
mport		kuč tsan78 masat	kua tsa7nő chin k	kua sa7ně fi7iš7 s masat		x seejs	
Import Multiple Files As		otravez vino venado	agui se vino agui	aca se viene este venadot.		DI THEM .	
lgit	Cel-G		Protocology (1997)			have the	
Mama de "L.	6 mas		y 7787 % gu		ti mapit no.	zuting	entonces ira kyli niju7
Trad Mama do "L	snadol,		y sate tigror,		venadote n	pare	entonses mira armadilij
Filts					luu	-	
Trad-Fain					116		
Jafio 1							

process is to export the ELAN transcription to an Excel file. The Excel file will become the basis for all of the analyses. This helps investigators to keep all of the analyses on the same Excel worksheet. Follow these steps to transfer the ELAN transcription to Excel:

1. Click on the File tab and then scroll down to the Export As tab and click on the Tab-Delimited Text option.

2. This action opens a window for exporting the transcription. Check that all of the tiers have been marked. (There is no need to export the default tier.) Then click ok at the bottom of the window.

X	🗶 Export as tab-delimited text	- 0 2
Elle Edit Annotation Jier Type Search View :	Export tier(s) as tab-delimited text	
Grid Text Sublities Lexicon Comments	Select ters	
UNI UNIX UNIX	Providence by Types' (b) Participants (b) Anexateurs (b) Languages Total Total	10 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

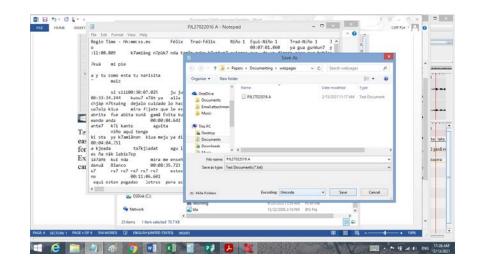
This action opens a third window that allows you to save the text file in the folder and format you wish to use. Be careful to use the UTF-8 encoding format. Select the destination folder where you wish to save the file and click save.

X	💐 Export as tab-delimited text	
Elle Edit Annotation Tier Type Search Yew i	Export tier(s) as tab-delimited text	
Grid Text Sublides Lexicon Comments In Solution Image: Solution of the soluti	Steric Gers By There Names By Types By Participants By Annotators By Languages Strong Strong Strong Stab-delimited text Serve Str. C webpages	01 11 2000 00 1000 00 1000 00 000 0
defend Cardina Cardina Cardina Transit Transit	Tile tjame P.A.27022016 A File Tormat (*.tro)	0 14 14 14 0 14 16 14 16 10 14 16 14 16 10 14 16 14 16 10 14 16 16 16 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10 10
Mans di A. <u>197</u> in ng TraoMans di A. 198 gg fan TraoMans di A. TraoFin TraoFin Cashologi	Exceding: UTT a Save Cancel Include the annotative M Include discription from the controlled vocatsilary Include discription from the control of the control o	rea musicy sizes in galaxies introduced in the second seco
Trad-bane 1	OK Close	

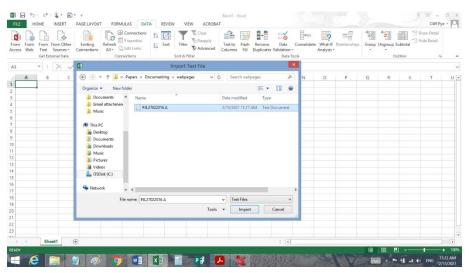
You should then find the resulting text file in the destination folder.

×		ELAN 5.1 - PIL27022016	A.eaf	- 5	X		
Eie Edit Annotation Tier Type Search		webpages					
Grid Text Subtiles Lexicon Con	File Home Share Vi	TW			^ O		
10 15 04 203 14 14 14 74 -H > >+ >F >F	Copy Patte shortout	More Copy Defete Rename folder	Properties Conn	Select all			
	🛞 🕣 * 🕈 📕 « Documer	ts + Papers + Documenting + webpages	v C Searc	h webpages	p		
PJL27022016 A. 👻 1.44.000 00 14.46.0			periodicinal Produced			0 00.15/02.000	00:15:04.000
	🚖 Favorites	Name	Date modified	Туре	See . *		
	Desktop	CC-by-sa	6/11/2020 4:16 PM	PNG File			
	Downloads	Documenting Language final	6/8/2020 9-47 AM	Microsoft Word D		1.	
1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1	32 Recent places	di equipment	7/4/2020 11:48 AM	Microsoft Word D		m bears	a a service a service
		Ma equipment	7/4/2020 12:07 PM	HTM File		+	
	CineDrive	Califavicon	6/11/2020 3:56 PM	GIF Film			
	🎍 Documents	favicon 1	6/11/2020 3:56 PM	lcon .			
default 000 00-14-46.0	🈹 Email attachments	tin Funding	6/29/2020 10:49 AM	Microsoft Word D		0 00-15-02.000	00:15:04.000
piy gua 6 kataka	JE Music	min funding	6/29/2020 11:49 AM	HTM File		ganā	o kataca kataka taka
an	Construction.	importance	6/27/2020 4:35 PM	Microsoft Word D.,			
Equi-JL ply gua ti ganãu,	💭 This PC	all importance	6/28/2020 4:45 PM	HTM File		ganā	u ganāu matū ganā m
tiene miedo tod	he Desktop	C index	6/9/2020 12:08 PM	Microsoft Word D		cabe	za calavera calavera
Tradul	Documents	nia index	7/4/2020 12:58 PM	HTM File			
Mamà de J. iu il	Downloads	a) mystyle	1/10/2009 12:15 PM	Cascading Style 5		R ti ganãu (xi7figA
(end	Music	PJL27022016 A	2/13/2021 11:17 AM	Test Document		I see see 1	su nar
Trad-Mamá de JL 12a dig	😩 Pictures	Children Children	6/26/2028 10:20 AM	Microsoft Word D		diga su ca.	su nat
Falls	🗃 Videos	all recording	6/26/2020 11:29 AM	HTM File		ilk7 k]	
198	SDisk (Ci)	C Returning	6/29/2020 10:49 AM	Microsoft Word D			
Trad-Felix		Min returning	6/29/2020 11:59 AM	HTM File		e drie	
	📬 Network	🔁 tža	12/22/2008 2:19 PM	JPG File	~		
Niño 1 Dis		٢			>		
Equi-fillio 1	25 items 1 item selected 70.7 KI	í			御日		
Trad-tillion		LA1					

Text files do not have special characters that specify font types and style types so text files are easy to process with computer programs. The text file was exported from ELAN in the UTF-8 format. This format needs to be changed to the Unicode format before importing the file to Excel. Open the text file in notepad. If you do not have the notepad app on your computer you can download the notepad++ app from the internet. Select the Unicode format and save the file.



Once the file is saved in the Unicode format, it is ready to import to Excel. Open a blank worksheet in Excel, click on the Data tab at the top of the screen, look for the Get External Data options, and click the From Text option. This action opens a window that lets you find the text file that you exported from ELAN. Click the Import tab at the bottom of the window and follow the instructions.



Select the delimited option and click Finish. This action should produce an Excel worksheet with separate columns that correspond to each tier in the ELAN transcription. The tiers for each participant are grouped together. The rows in the worksheet show the beginning time for each utterance by each participant.

		Sort & Fi Filter - Se	THE FUL		a second s	Formatting	#E Merge & Cente	100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100 - 100	x y + ⊞ + <u>⊅</u>	Format Painter	aute ,
		Editing		Cells	Styles	G Number G	Alignment		Font	Dipboard IS	
									√ fx	• X v	£
i	1	н	G			£	D	c		A	
01	Trad-Niño :	Equi-Niño 1	Niño 1		Trad-Félix	Félix	Trad-JL	Equi-JL	ms A	in Time - hhommoss.ms	849
					sili	juuu			24.1	01:24.1	
					si sii pueblo	ju juu kam7us			34.4	02:34.4	
				000	es uno pueblo	ss nda kam7us mē			36.9	02:36.5	
					no	mě			39.2	02:39.2	
					si	jâ			40.5	02:40.5	
					si sii	ja jua			42.0	02:42.0	
					bueno	mjau			43.7	02:43.3	
				abla mas	mas otra vez ha	mas kara k7unèun7at mås			45.6	02:45.6	
					venga Luis	k7un7ēt Luis			47.9	02:47.5	
					se callo	gatsu7	calzon binco	dankjúy7 danuã	05.3 tucha	03:05.3	
				el	se cayo un pape	gatsu7 nda nk7kx			06.7	03:06.7	
					como esta	pik xiñ7chja7			21.6	03:21.6	
					casas podrida	nt7üs makã			57.6	03:57.6	
					588	juou			59.7	03:59.7	
					si sii	ju juu			15.5	04:15.5	
					oso quiote	oso sampu			17.9	04:17.5	
				sigue hablale	sigue hablale si	siga riajaut siga riajaut			54.6	05:54.6	
				ndo	que siga hablan	ke siga gunèun7				05:56.8	
					ya va hablat	ya gua gunèun7				07:01.1	
				iera	habla como qui	nèun7 komo kiera				07:04.3	
					si sii	ju juu			09.9	07:09.5	
					hablale hablale	riajaut riajaut				07:11.4	
					dite eine ereche	alles to WOW das a Whatasaa			····	· Sheet1	

Save the worksheet with the same name as the ELAN file in order to identify the source of the data in the Excel file.

The Excel provides sorting and filtering tools that help in data analysis. The worksheet should preserve any special characters that were used in the transcription. The first step in analysis is to create separate worksheets for each participant. Start by selecting the first four columns, which in this example show the Begin Time, JL's utterances, the adult equivalences, and the translations.

aste	X Cut Calibri - 11 Copy - B I U - ⊡ - 2	• <u>∆</u> •≡≡≡ e	Y + ■ Wrap Text E #E Merge & Center	Formatting *	Table * Styles * * * *	E Fill	Sort & F ar * Filter * S	
	lipboard Is Font		Abgnment	G Number G	Styles Cells		Editing	
1	👻 i 📉 🧹 🎵 Begin T	ime - hh:mm:ss.ms						
	A B	c	D	E	F	G	н	10
Beg	in Time - hh:mm:ss.ms_JL	Equi-JL	Trad-JL	Félix	Trad-Félix	Niño 1	Equi-Niño 1	Trad-Niño 1
	01:24.1			Juuu	sill			
	02:34.4			ju juu kam7us	si sii pueblo			
	02:36.9			tes nda kam7us mě	es uno pueblo no			
	02:39.2			mē	no			
	02:40.5			18	si			
	02:42.0			lju juu	si sii			
	02:43.7			mjau	bueno			
	02:45.6			mas kara k7unèun7at mấs	mas otra vez habla mas			
	02:47.9			k7un7ët Luis	venga Luis			
	03:05.3 tucha	dankjúy7 danuã	calzon blnco	gatsu7	se callo			
	03:06.7			gatsu7 nda nk7ix	se cayo un papel			
	03:21.6			pik xiñ7chja7	como esta			
	03:57.6			nt7ùs makā	casas podrida			
	03:59.7			ไม่มนม	siii			
	04:15.5			ju juu	sisii			
	04:17.9			ioso sampu	oso quiote			
	05:54.6			siga riajaut siga riajaut	sigue hablale sigue hablale			
	05:56.8			ke siga gunèun7	que siga hablando			
	07:01.1			ya gua gunèun7	ya va hablat			
	07:04.3			nèun7 komo klera	habla como quiera			
	07:09.9			ju juu	si sii			
	07:11.4			riajaut riajaut	hablale hablale			
	► PJL27022016 R (+)			Laborator Mattalase - Minteres	dife and a sector factor for a sector of a			

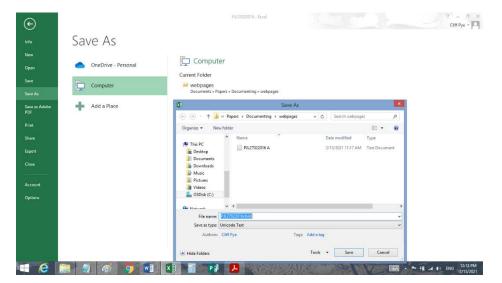
Copy these four columns to a new worksheet in the same file. Add an additional column to record which part of the recording the transcription comes from.

Cut Ella Cop Paste Clipboar	y - B I U - ⊡ - ∆ - A -	= = 			Conditional For Formatting * T		Insert Delete Format	∑ AutoSum * A Till * Sort & Fil Clear * Filter * Se Editing	
1	• : $\times \checkmark f_X$ Part								
A	B C	D	E	F	G H	3	JK	L M N	0
Part	Begin Time - hh:mm:ss.ms JL	Equi-JL	Trad-JL						
A	01:24.1								
A	02:34.4								
A	02:36.9								
A	02:39.2								
A	02:40.5								
A	02:42.0								
A	02:43.7								
A	02:45.6								
A	02:47.9								
A	03:05.3 tucha	dankjúy7 danuã	calzon binco						
A	03:06.7								
A	03:21.6								
A	03:57.6								
Α	03:59.7								
A	04:15.5								
A	04:17.9								
A	05:54.6								
A	05:56.8								
A	07:01.1								
A	07:04.3								
A	07:09.9								
A	07:11.4								
	07020.2								

The resulting worksheet contains all of the child's utterances with their interpretations and translations. The beginning times make it possible to find each utterance in the ELAN transcription. The analyses of the children's utterances start from this worksheet.

Lexical Analysis

The first step in lexical analysis is to produce a lexical concordance for the children's words. A concordance provides the utterance contexts for a child's production of each word in the recording. It is best to use the adult targets as the concordance keys in order to group together all of the child's attempts at each adult word. Because it is easier for computer programs to process text files, it is necessary to save the Excel file as a Unicode text file. Go to the File tab in Excel and select the Save As option. Excel will open a window in which you can select a folder for the text file as well as the file format. Select the Unicode Text format and click Save.



Excel will open a window that says the selected file type does not support multiple sheets. This is not a problem since we are only interested in saving the sheet with the child's data. Click OK.

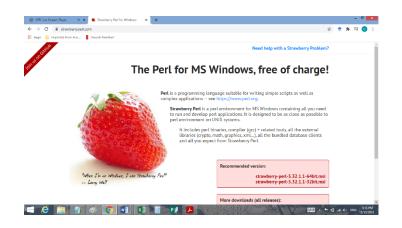
aste Gipboar	mat Painter B I U + 🗄 + 👌			Wrap Text Merge & Center - r		+ 0 00 Con Form	ditional Form satting = Ta Style	ble * Styles *	Insert	Cells	∑ Auto ↓ Fill •	s	ort & Find & ilter - Select		
41	- : $\times \checkmark f_x$ Part														
A	B	С	D	E	F	G	н	1	1	К	L	M	N	0	
	Begin Time - hh:mm:ss.ms JL		Equi-JL	Trad-JL											
A	01:24.1														
A	02:34.4														
A	02:36.9														
A	02:39.2			1		64		-		×					
	02:40.5				Microsoft Exe	:ei									
A	02:42.0		The column diffe time door	and successful and the											
A	02:42.0 02:43.7	A	The selected file type does												
A A	02:42.0 02:43.7 02:45.6	A	• To save only the active sh	eet, dick OK.	oks that contain m	utiple sheets.	boose a file t	ype that suppo	arts multipl						
A A A	02:42.0 02:43.7 02:45.6 02:47.9	A		eet, dick OK.	oks that contain m	ultiple sheets. ne for each, or (choose a file t	ype that suppo	orts multipl						
A A A A	02:42.0 02:43.7 02:45.6 02:47.9 03:05:3 tucha	A	• To save only the active sh	eet, dick OK.	oks that contain m	utiple sheets.	choose a file t	ype that suppo	orts multipl						
A A A A A	02:42.0 02:43.7 02:45.6 02:47.9 03:05.3 tucha 03:06.7	A	• To save only the active sh	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						
A A A A A A	02:42.0 02:43.7 02:45.6 02:47.9 03:05.3 tucha 03:06.7 03:21.6	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						
A A A A A A A A A	02:42.0 02:43.7 02:45.6 02:47.9 03:05.3 tucha 03:06.7	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						
A A A A A A A A A	02:42.0 02:43.7 02:45.6 02:47.9 03:05.3 tucha 03:05.7 03:21.6 03:27.6	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						
A A A A A A A A A A A	02:42.0 02:43.7 02:45.6 02:47.9 03:05.3 tucha 03:06.7 03:21.6 03:57.8 03:57.8	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (choose a file t	ype that suppo	orts multipl						
A A A A A A A A A A A A A A	0242,0 0243,7 0245,6 0247,9 0305,3 tucha 0305,7 03:21,6 03:57,6 03:57,6 03:59,7 04:15,5	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (choose a file t	ype that suppo	orts multipl						
A A A	02242.0 02243.7 02245.6 0235.3 0336.3 tucha 0336.7 0337.6 0337.6 0337.6 0339.7 0415.5 043.7.5	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						
A A A A A A A A A A A A A A A	02420 0243.7 0245.6 0247.9 0336.5 tucha 0336.5 tucha 0337.6 0337.6 0359.7 0415.5 0417.9 0554.6	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						
A A A A A A A A A A A A A A A A A A A	02242.0 02243.7 02345.6 02347.9 03356.7 03321.6 0337.6 0337.6 03359.7 0415.5 0415.5 0417.9 0534.6 0534.6	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (choose a file t	ype that suppo	orts multipl						
A A A A A A A A A A A A A A A A A A A	0242.0 0243.7 0245.6 0247.9 0350.5 tucha 0335.7 0321.6 0337.8 0339.7 0415.5 0417.9 0554.6 0554.6 0554.8	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (choose a file t	ype that suppo	orts multipl						
A A A A A A A A A A A A A A A A A A A	0242.0 0243.7 0243.7 0245.6 0357.6 0357.6 0357.6 0357.6 0357.6 0359.7 0415.5 0415.5 0417.9 0554.6 0554.6 070.1 070.1	A	 To save only the active sh To save all sheets, save th 	eet, click OK. em individually using	oks that contain m	ultiple sheets. ne for each, or (thoose a file t	ype that suppo	orts multipl						

This action should produce the Unicode text file containing the child's utterances. The text file will have five tab-delimited columns that correspond to the columns in the Excel file. The first line in the text file will have the column headers shown in the first row of the Excel file. Save the file in the UTF-8 format in order to use it in the analysis programs.

	Sort & Find & Filter - Select	+ 🖑 Clear -	Conditional Format as Cell Formatting * Table * Styles *	General ter + \$ + % + \$		ibni •[11 •]A I¥ • ⊞ • ⊉ •	Copy -	
	ng	- D Edda	ild - Notepad	РЛ.27022016	a	Font	boerd fa	Cipb
					File Edit Format View Help	√ f _X Part	• X	41
0	N	л м.	ui-3L Trad-3L	s IL	Part Begin Time - hh:mm:ss.m			A
	19				A 01:24.1 A 02:34.4	hommoss.ms JL	Reals Time hi	Part
					A 02:36.9	01:24.1	Bellin must - re	A
	×		Save As		A 82:39.2	02:34.4		A
		v & Search webpages ,0	Documenting > webpages v	(b) (b) (c) (t) 🚺	A 02:40.5	02:36.9		A
	4	A C Search medbades To	ocumenting + weopages •		A 02:42.0	02:39.2		A
			der	Organize - New	A 02:43.7	02:40.5		A
					A 02:45.6	02:42.0		A
		Date modified Type	Name	This PC	A 02:47.9 A 03:05.3 tucha dankjū	02:43.7		A
	Doci	2/13/2021 11:17 AM Test D	PJL27022016 A	Desktop	A 03:06.7	02:45.6		A
	loci	2/13/2021 3:10 PM Text D	PJL27022016child		A 03:21.6	02:47.9		A
			1.000	Documents	A 03:57.6	03:05.3 tucha		A
				Downloads	A 03:59.7	03:06.7		A
	>		· <	Music	A 04:15.5	03:21.6		A
	-		7023016-bad	File name	A 84:17.9	03:57.6		A
	100				A 05:54.6	03:59.7		A
	¥		Documents (*.txt)	Save as type: 1	A 05:56.8 A 07:01.1	04:15.5		A
					A 07:01.1	04:17.9		A
					A 07:09.9	05:54.6		A
	1	V Save Cancel	Encoding: UTF-8	120000000000	A 07:11.4	05:56.8		A
	10	Carcel	uning vira	Hide Folders	A 07:23.3	07:01.1		A
					A 07:24.9	07:04.3		A
		~			A 08:19.9	07:09.9		A
		•			¢	07:11.4		Α.
			1. C. 1.		(*)	6 PJL27022016child	P.8.27022016	
			1.141		۲	 Price reg2016child 	PA27022016	4 14

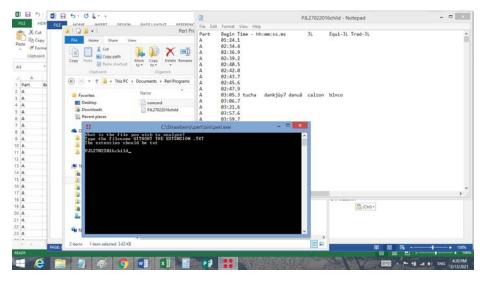
Perl Analysis Programs

It is easy to produce a lexical concordance with programming languages that are specialized for text analysis. Perl is an older one of these languages, while other programming languages like Python will also work. A computer program is a list of instructions that the program interpreter translates into computer instructions. In order to run Perl programs you must first download and install a Perl interpreter. The Strawberry Perl interpreter works on both PCs and Macs so it is a good choice. Strawberry Perl can be downloaded from the webpage http://strawberryperl.com>.



In order for computer programs to work, the programmer has to tell the program where to find the data file. The easiest way to handle this problem is to open a new folder and label it Perl Programs. This folder will serve as the place to use all of the analysis programs. You will also need to copy the child's UTF-8 file to this folder for the analysis. The lexical concordance program concord.pl can be downloaded from the webpage < http://pyersqr/minimal>.

With the concord.pl file and the data file, e.g. PJL27021016child saved in the Perl Programs folder, start the program by double clicking on the concord.pl file. A program window will open and ask you to type in the name of the data file, in this case PJL27021016child. Type in the filename and click the Enter key on the computer's keyboard.



When the program runs successfully, the program window will close and the program will output the concordance file in the Perl Programs folder with name of the datafile, but with a new extension '.con'. The concordance file contains the results in tab-delimited columns that can be imported into the Excel worksheet. In order to preserve any special characters, open the concordance file in notepad and save it in the Unicode format.

A Pr X	Hene Share Vec Image: Share X Cat Partir Copy path Partir Image: Copy path Partir this thirt out Copy back Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path Image: Copy path	Nore Copy Delete Rev Copanies Documents + PetiProgram	7a 71ε 7mεjε a	A 30:45. A 27:24. A 27:34. A 03:17. A 16:17.		71z no se enoja hombre	ng mà m7jèng ros los dos
Pa 🗙	Desktop			1			
	Desktop				Save As		×
A 3				⊕ ⊕ = †	= Docume > Perl Programs v (Search Parl Programs	P
A 3		P/L27022016child . co	a7a	0.0		•	
	Recent places	PR.27022016child	88	Organize • No	lew folder	10 ×	•
			abejent		* Name	Date modified	Type
	OneDrive		achi	1 This PC	PJL27022016child	2/18/2021 4:05 PM	Text Doc.
	Documents			Desktop			
2 P 2 P 2 P 2 P 2 P 2 P 2 P 2 P 2 P 2 P	Email attachments Music			Downloads			'p ac
4	Mubic			1 Music	v (> p ac
5 4	This PC						Pp ac
6	Desktop				PR.27022016child . con		y p ac
	Documents		aja	Save as type:	E Test Documents (*.tst)		v pac
9	Downloads		- 1-				
10	Music		<				
11	Videos			Hide Folders	Encoding: Unicode	V Save Cano	cel
	OSDisk (C)				and a second second second		1.1
14							
	Network						
		¢			>		
+ PAGE 3 iter	ns 1 item selected 78.2 KB				E AL AND		

The lexical concordance is now ready to be imported into the Excel file. Open a new worksheet in the Excel file, rename it lexical, and click on the data tab in order to import the text file. Look for your Perl Programs folder and change the format at the bottom of the window from txt files to all files. You should then see the concordance file that the program produced. Click Import.

com From From From Other ccess Web Text Sources* Get External Data	Connections All + Li Edit Links A+	Text to Flash Remove Data Advanced ta Fitter Data	Consolidate	What-W Aslations Analysis *	Group Ungroup	Subtotal	
u • i 🛛 🗸 🗸	C) imp	ort Text File	×				
A B C	🛞 💮 👻 🕆 🎍 🕫 Documents 🔹 Perl Programs	v 🖒 Search Perl Programs J	Ø N	0 P	Q R	5	T U
	Organize - New folder	E • 🖬 🕯					
	Conced Concent Concen	2/1/201322784 Pet program // 2/11/20214229M File 2/11/20214039M Test Document					
	File name: PR.27022016child . co	n v All Files . Tools v Import Cancel					

The Excel program will ask again if the file is delimited. Click the Finish tab, and the concordance will be imported to the new worksheet.

E La La	es Web Tr	Trues Other of Sources Ethional Data	Lostray Correction	Connections 24 Image: Connections 24 Image: Connections Image: Connections 24 Image: Connections Image: Connections	Teatro Flach Remove Data Consolid Colorna Fil Dapleater Validation * Data Inoti	Analysis + + +	Substal
64	+	1 X V	fx .				
		8	c 0	4 /	G.	н	1.1.1
13	Adult Word	Child Word	category part	t time child utterances	adult utterances	translation	
	7#	7#	A	11.51.4 je 7e mě	ja 7a mē	no no no	
		fiziāz	A	30345.0 jaža řízlůz mů mjöng mů mjeng	ja za fiziāz mā miņēng mā miņēng	no este van ajugar van a jugar	
2	714		A .	27/28.8 tule	5 74	hombresote	
		1£	A	27:34.9 ni natsutng k	ni natsuong 71c	no se onoja hombro	
5 }	7mcje	mjž	A	03:17.4 kamione mjë nuye	komiune 7mcjenuye	carros los dos	
			A	15.17.8 ti kemiutodo	te kami un 7meje	carrote todo	
	8	8	A	13:14.3 atajūž	a tai/jū/t	lo nombro	
2		8	A	03:35.7 a sapujt	a saput	sapcte	
0		919	۸	15:35.8 ana	a mõ	ano	
1			A	24:05.4 a skuru	a kempiujun7	a oscuro	
2			A	20:10:9 ljižíkh e si	ljíáňch a sil	hilos a asii	
	a7a	14	A	32:55.9 ja 7a mē	a7a mð	no no	
4		jus	A	81:48:0 ju juu	ja aa	51.51	
	abcjient	abent	Α	27:42.7 lifkiuEabent	rañ7kiut absjent	lumbres traian	
	achi	achi	A	32:03.8 achi bulo achi bulu	achi nmc7p	achi burro	
7		achi	Α.	32.05.0 achi bulu	echi nmc7p	achi burro	
8		achi	A	34:15.9 achi bulo	achi ninc?p	achi bunito	
9		achi	A	34:18.2 achi bulo achi bulo achi bulo be te	achi nmc/p achi nmc/p achi nmc/p biyy tê7	achi burro achi burro achi burro asi d	
0		achi	4	34:18.2 achi bulo achi bulo achi bulo be te	achi nmɛʔp achi nmɛʔp achi nmɛʔp bɨjɨy têʔ	achi burre achi burro achi burro asi d	
1		achi	A	34:19.2 achi bulo achi bulo achi bulo be te	əchi nmɛʔp əchi nmɛʔp əchi nmɛʔp bijiy têʔ	achi burro achi burro achi burro asi d	
2		achi	A	34:19.9 achi bulo achi bulo achi bulo be te	echi nmc7p echi nmc7p echi nmc7p bijiy tê7	achi burro achi burro achi burro asi d	
1		achi	A .	34:19.9 achi bulo achi bulo achi bulo be te	echi nme7p echi nme7p echi nme7p bijiy tê7	achi burro achi burro achi burro asi d	

The concordance sheet contains 8 columns, in this case labeled: Adult Word, Child Word, category, part, time, child utterances, adult utterances and translation. The Adult Word column contains all of the words that the program found in the adult utterances column and shows all of the utterances that contained that word. In most cases, the child's line will contain the child's attempt at producing the adult word, but in some cases, the child might have omitted the adult word. The concordance program extracts a word from the child's utterances every time it extracts a word from the adult utterance. When there is not a one-to-one match between the number of words in child's utterances and the adult's utterances the program extracts a mismatch. Examples of these mismatches can be seen on rows 3 and 4 in the example. In row 3 the program matches the adult word 7a with the child's word ñ7iã7. This mismatch occurred because the adult utterance begins with two words—ja 7a that correspond to the single word ja7a in the child's utterance. Such mismatches could be corrected in the original transcription, but this would take time. It is much easier to fix such mismatches at the analysis stage because they are easier to spot.

The variation between the child's production tokens for each adult target contains information about the stability of the child's productions. This information can be used to address the research question of what factors affect the child's productions. Children produce some words more faithfully than others, which reflect the child's ability to produce specific sounds, syllables and words. Comparisons between different children will reveal which factors the children have in common and which productions are specific to individual children.

Selecting a Representative Phonetic Type

We assume that children store a single representation for each word in their lexicon. David Ingram (1989) published a method for selecting a phonetic token from the child's productions for each word as a representative phonetic type. The phonetic type is chosen on the basis of the features that it shares with the other tokens. In most cases, the phonetic type is the token that the child produce more frequently. If frequency does not yield a phonetic type, then the token that shares the most phonetic features is chosen as the phonetic type. It is easy to use the lexical concordance to select the child's representative type for each word. The first step in this process, as in all subsequent analyses, is to copy the child's lexicon to a new worksheet. The first rule of analysis is to preserve a copy of each step in the process so that it will be possible to recover an earlier step in the process if a mistake is discovered in the analysis. Copy the child's lexicon to a new sheet and label it Types.

	∑ AutoSum * A ↓ Fill * Z ✓ Clear * Filter * S		Conditional Format as Cell Formatting * Table * Styles	General + \$ + % + % .23	▲·▲· 王 王 王 任 把 E Merge & Center ·		Painter 8 I	te Format	Par ·
	Editing	Cells	Styles	Number G	G Alipsment C	Font		Clipboard	
						∫x 7a	XV		12
1	4.		6		1 F	6	A	d	
		ranslation		adult utterances	t time child utterances	Child Word	Adult Word		
		o no no		ja 7a mē	11:51.4 ja 7a mě	7a	7a	Į	
	ugar	o este van ajugar van a ji		ja 7a ñ7iā7 mā m7jēng	30:45.6 ja7a ñ7iā7 må mjöng mā mjong	A	ñ7i37		
		ombresote		ti 7%	27:24.1 tule	A		71£	
		o se enoja hombre		ni natsueng 71e	27:34.5 ni natsueng le	A	lε		
		arros los dos		kamiune 7mEj£ nuye	03:17.4 kamione mjä nuye	A.	mjil	7maja	
		arrote todo		ti kamiun 7meje	16:17.8 ti kamiutodo	A			
		o nombro		a ta7jū7t	13:14.3 a tajū7t	A		a	
		apote		a saput a mě	03:36.7 a sapujt	A	4		
		no			16:35.8 ana 24:06.4 a shuru	A	ana		
		oscuro ilos a asii		a kampiujun7 Isiāřich a sil	24:06.4 a skuru 26:10.9 liiäñch a si				
		nos a asin o no		ijianch a sii a7a mē	26:10.9 ijianon a si 32:55.9 ja 7a mē	A	a ja	a7a	
		o no I sil		ja aa	32:55:5 ja 7a me 31:48.0 ju juu	A	ja juu	a/a aa	
		ambres traian		ja aa rañ7kiui: abejent	27:42.7 liňkiuj abent	A	abent	aa absjent	
		chi butto		achi nmt7p	32:03.8 achi bulo achi bulu		achi	achi	
		chi burro		achi nmi?p	32:06.0 achi bulu		achi	acini .	
		achi burrito		achi nmi7p	34:16.9 achi bulo		achi		
	i burro así dice	chi burro achi burro ach		achi nmc7p achi nmc7	34:18.2 achi bulo achi bulo achi bulo be te	A	achi		
	hi burro asi dice	chi burro achi burro achi		achi nmc7p achi nmc7	34:18.2 achi bulo achi bulo achi bulo be te		achi		
	hi burro asi dice	chi burro achi burro achi		achi nmi?p achi nmi?	34:18.2 achi bulo achi bulo achi bulo be te	A	achi		
	hi burro esi dice	chi burro achi burro ach		achi nmc7p achi nmc7	34:19.5 achi bule achi bule achi bule be te	A	achi		
	ni burro asi dice	ichi burro achi burro ach	p achi nmɛ7p bijiy tê7	achi nmc7p achi nmc7	34:19.9 achi bulo achi bulo achi bulo be te	A	achi		
					Schild Lesicon Types	P.8.27022016	PA27022016		

As well as identifying a representative phonetic type, it is also helpful to identify the word's part of speech. It will then be possible to how frequent each syntactic category appears in the child's utterances. I add an abbreviation in the category column that identifies the word's lexical category. I move the adult word and the category symbol to identify the representative phonetic type. I use the following list of category symbols, but others can be added as needed.

Lexical category symbols

A adjectives, adverbsO onomatopoeiaC common nounsP pronounsD demonstrativesQ wh-words, yes, noE exclamationsT transitive verbsI intransitive verbsX uninterpreted

N names

The following example shows the worksheet with the representative types and lexical category information. The selected cells in this example show that I moved the adult word 71 ϵ to the fifth row in order to select the child's token 1 ϵ as the representative type for this word. The lexical category for the word is a common noun so I added a C in the category column.

	Sort & Filter * Set	∑ Autofium ↓ Fill * ≪ Clear *	Delete Format			Conditional Format as contacting * Table* Shirs	General +) \$ - 96 + 52 23 Number 52	Wrap Test		- A	* [1] = + Fact	г <u>и</u> •	Tainter R J	te di Cus le Copy -	Pat
											74	/ fx	IX 4		Å.
i i		1						A B C D I I							
				anslation	24		adult utterances	Child Word category part time child utterances							1
				ono no	no		ja 7a mē		a mê	A	E	7a	7a		
		igar	n ajugar van a ju	este var	no	iā m7jēng	ja 7a ñ7lā7 mā m7jêng	mã mieng	A		ñ7@7				
			e	mbresot	hor		ti 71s	27:24.3 tule							
			a hombre	se enoja	ng		ni matsueng 7%	27:34.5 ni natsuarg la				c	la.	7ia .	1
			las	eros los d	car		kamiune 7mejenuye	03:17.4 kamione mjž nuve				A	mjž	7meje	
			la .	rrote tod	car		ti kamilun 7mcjic	16:17.5 ti kamiutodo				s	todo	todo	
				nombro	lo r		a taījū?t	13:14.1 a tajū7t				1			
				pcte	sap		a saput	03:36.7 a sapujit					a		
				10	an		a mē	16:35.8 ana					ana		2
				oscure	20		a kamplujun7	24:06.4 a skuru					a		r j
				los a asii	hile		ljiäñch a sil	26:20.9 ljižřich a si					4		
				ono	00		a7a mê		a mē	32:55.	A	E	ja	a7a	2
				516	3i 5		ja aa		uu	31:48.	A		juu	88	I.
			lan	mbres tra	lun		rañ 7kiut abojent		tiut abent	27:42.	A.	T	abent	abcjant	
				hibutto	ach		achinme7p		i bulo achi bulu	32:03.	A	С	achi	achi	s
				hibuno	ach		achinmc7p		i bulu	32:06.	A		achi		1
			10	chi burri			achinm(7p	34:36.9 achi bulo					achi		1
	č	i burro asi dice	achi burro achi	thi burro	7 30	chi nmɛ7p bijiy tê7	achi nmc7p achi nmc7p	34:18.2 achi bulo achi bulo achi bulo be te					achi		2
	6	i burro así dice	achi burro ach	thi burro	7 at	chi nme7p beyy tê7	achinme7p achinme7p	34:38.2 achi bulo achi bulo achi bulo be te					echi		5
		i burro asi dice	achi burro ach	thi burro			achinne?pachinne?p	34:18.2 achi bulo achi bulo achi bulo be te					achi		1
			achi burro ach				achinme7p achinme7p	34:19.9 achi bulo achi bulo achi bulo be te					achi .		1
			achi burro ach				achi nmc7p achi nmc7p		i bulo achi bulo a		٨		achi		1
		it has not all as		-h-1.h	9	AL		A16.7+6+++	con Types		27022016		P.R.27022015		

Once the representative phonetic types are selected, the custom sorting option in Excel can be used to group together the lexical types that belong to each lexical category. The sorting should be done by copying the sheet to a new worksheet and sorting the words in the new sheet as shown below. The selected cells in this example show the adult word 71ϵ 'man' is the first common noun in the worksheet.

B Copy		Calibs - 1		• A' A'	H = 🞆	æ,	Wrap Test	General +		•			9	뢂	*		∑ AutoSu Fil+	ZY			
Parte Format		IU -	8 •)	Q · 4 ·	0.0.5	包积	III Merge & Center +	5 - % .	· 17.		onditional smatting -	bonal Format as tting - Table -	Cell Styles -	Insert	Delete	Format	Cear -		& Find &		
Clipboard	- 6	To	et			Alige	ment 15	- Not	ber	4		Styles			Cells			Editing			
$f_{ADS} \rightarrow f_X \checkmark f_X \mid \pi_0$																					
	· · · · · · · · · · · · · · · · · · ·																				
A	8	C	D	E 00.31.6 se	- 1 m 1			6								H		-	3		
56 sasua7	54567 5470	Α.	A								gris										
57 så7p 58 skua	sa7p skua	A	A	11:07.8 gu	asa7p 7 skua n7gul				is sā?p mas deirato 7 skua n7puà me voy pora alla												
50 sxua 59 sua7	soa7	A .	Å	00:09.2 so	5075808	uv8na				az		es pils									
50 ta7kiii5t	takili	A .	2	02:20.4 ja	ja jao lañj	0:08-ot t	- the set				sii pasar	ralar									
51 tů	to	A	A	23:18.2 to	tů chaléje			24.			tonses :		alizana	100							
52 xiñkji#7ts	xiñchu/7s	A	Â	08:28.0 xil		U.S.		sifiki#7ts	ing a string	ios.				orado	PE SHILD	ercone	pito				
53 Ak7iy	ki7	A	A		nkjea mja ki	7		k7i nkjuž	milku k717	5k7iv			el	coneito	bies	Nouito	13 - C				
54 Akiijell7	k17/6	4	4	08:35.7 kill				Skille17						ede							
55 +7xi378u	xia7au	A	A	27:52.0 ch	27:52.0 chi xia7au må gatiñ				78 u mẫ a:	a7tiñ			ah	ora otro	va entr	egar					
56 71c	le	C	A	27:34.9 ni	7:34.9 ni natsueng le				g 71c				no	se enoi	a hom	bre					
57 achi	achi	c	A	32:03.6 eci	2:03.8 echi bulo echi bulu				p					hi burro							
50 bi7kjiuàk	cuchiy	c	A	18:55.3 14	18:55.3 ti-cuchiy				k				m	machetote							
59 chijijiñ	chji/2iň	с	A	28:37.4 chji7iň				chysjeň													
70 dagua7a	dagua7a	C	A	21:15.6 dagua7a				dagua7a													
71 dankjùy7	dankúy7	C	A		a ti dankûy7			kua ti dankjūy7					ac	aca pantalonsote							
72 deù	dò	C	A	19:36.1 dò				drù						rro							
73 ga7m7u7	mu7	с	A	31:13.0 mu7				ga7m7u7						tortuga							
74 gameng	kamin	с	A	21:54.4 ch	ch7i game	ng					ardillita										
75 gamii7i	mi7i	с	Α.	21:02.1 mi		gam670						eja									
76 gamès	games	C	Α	25:03.3 ky		k7i gamits						añita									
77 ganãu	ganao	c	Α	09:20.2 ña				ñ7iã7 gan						ta cabez							
78 ganjäudat	ñjaudat	c	A	30:35.9 m	jau belidat m	a riyaud	at t7ueye mã njiãche	mjàu bah	¢at mā ga	injilud	lat nt7ua r	män7jäud		uchos se	t subier	on al ce	rro ben a co	merselos			
	PH27022016	P.8.270	02016	bid Lin	icon Typ	0	tegories (+)				1.6										1(*)

The number of lexical types that belong to different lexical categories such as intransitive and transitive verbs, etc. are easily counted in Excel. Excel also makes it possible to rapidly calculate the percentage of lexical types in each category. For examples of research on children's lexical categories see Gentner & Boroditsky (2001); Pye, Berthiaume & Pfeiler (2020).

Vowel Analysis

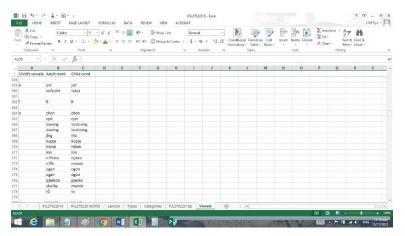
The representative types provide the data for a basic phonological analysis following Ingram's method for analyzing child phonology (1989). I wrote the perl program vowel.pl to extract the vowels from a child's phonetic types as the preliminary step to analyzing the accuracy of the child's vowel production as well as the substitution patterns that occurred. As always, the first

step in a new analysis is to copy the child's lexical types to a new worksheet. This step is necessary when the analysis will not include all of the words that the child produced. The idea is to only include the child's words that correspond to identifiable adult words. This step involves deleting the child's uninterpretable words as well as such categories as exclamations and onomatopoeia form. The example shows the sheet with these categories omitted.

Cut Cop Cop Cop Cop Cop Cop Cop Cop	mat Painter	Calibri B I	* 1 ⊻ * ⊞ * Font	1 ~ A'	≡	+		Viap Test Aerge & Cent	Gene er • \$ •				ormat as Ce Table - Styl pics		t Delete I	ormat	AutoSum Fill - Clear - E	Sort & Fi Filter * Se		
J248	• i]	× v	f_X																	
A	В	с	D	Ε	F	G	н	1.1	1	ĸ	ι	м	N	0	P	Q	R	S	т	
5 117	j#7	P	A	09:06.1	1 ti ganău	ji ti ganão	j tu cabea	rota tu												
7 kiau	kiao	P	A	02:30.7	7 kiao kjo	an kiau nk	ur to tamb	ién												
8 kiaugue7		P	A	19:02.9	9 bia7ang	a7 ian7gue	7 mandar	ios a nosotri	os											
9 กลกนี้	nanû	P	A	33:19.5	3 nanû ki:	su nanû ki	au yo viya	0												
	u	P	A					in a coparar												
uye	u	P	A	20:48.8	3 u libie7	se uye lab	icji ellos se	empalman												
	bijiy	Q	A					de esta peg	adito											
bijiy	bijiy	Q	A				oij burro as													
mä	mâ	Q	A	29:11.0) samje m	å sam7je	mâmaripos	ia va a traer	uno											
5 mè	millili	Q	A	35:03.0	i miiiim	mė	no													
6 mē	me	Q	Α	10:21.3	3 me bolt	ec më volt	ec no volte	10												
í mž	mž	Q	A	35:19.7	7 mž	mž no	no													
nip	ni	Q	Α	09:39.7	7 ni ju7u	nip ta7j	u7 no puec	lo												
nap	nep	Q	A	06:33.3	3 nep blu:	7 nep blu	7 que es e	250												
nik	nep	Q	Α		3 nep bio		ik cual se t													
płk	pak	Q	A	25:22.4	t mama p	aknàn pik	la mamá a	donde vas												
absjent	abent	т	A				2 a lumbre:													
bisjst	libiat	т	A					tita Ilanta II												
biè7ɛulat			A					a avientalo:	5											
	chalEng		A				ncle mont													
chi7i878ñ	ki7iä7äñk	т	A					enseñaste		ro										
cjot	cjot	т	A					su nombre												
da7juÊ7êr		T	A			n:da7juÊ7	êr lo levan	to												
<	PJL270	22016	PJL2702201	6child		Types	Categor		ology	÷		: •				II			1	

The vowel analysis program compares each adult word with the child's representative phonetic type. Once again, it is necessary to save this worksheet as a Unicode text document for the perl program. Open the Unicode text document in notepad and change its format to UTF-8. Copy this file to the Perl program folder and run the program vowel.pl by double clicking on the file. The program will produce a UTF-8 text file with all of the child's and adult words that contain each vowel.

The UTF-8 file needs to be converted back to the Unicode format before importing the results to excel. I added a new worksheet labeled vowels in Excel and imported the vowel data into this sheet.



As shown in this example, the vowel program outputs the results in three columns. The first column contains the target vowel, the second column contains the adult target word, while the third column shows the child's word. As shown in the first row, this is an analysis of the vowels that were in the child's words. In this analysis, the program found all of the child's words that had the target vowel. Note that the program outputs the child's word for each vowel so if the

word contains two or more instances of the same vowel, the word will be shown that number of times. In this example the child produced two words (tostosing, ogon) with two instances of /o/ so the program shows these words twice. This output allows the investigator to count how many phonetic types the child produced with each target vowel. This child produce /o/ 15 times in 13 phonetic types.

The program does not separate vowel strings in order to make it easier to separate monophthongs from diphthongs. The example shows the child's words with the vowel sequence [ii] at the top. The investigator can analyze such sequences as either successive monophthongs or diphthongs depending on the goal of the analysis.

The program outputs an analysis of the vowels that were in the adult target words separately from the analysis of the vowels in the child's words. There are often discrepancies between the vowels in adult words and the vowels that children produce in their own words. An analysis of the number of times the child produced the vowels in the adult words provides an indication of the accuracy of the child's vowel productions. A final determination of a child's vowel productivity will be based on the frequency of use and accuracy for each vowel.

Analyses of vowels in children's words are still rare. Examples of such analyses can be found in Bernhardt & Stemberger (1998), Fee (1991) and Levelt (1994).

References

- Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development: From the perspective of constraint-based nonlinear phonology*. New York: Academic Press.
- Gentner, D. & Boroditsky, L. (2001). Individuation, relativity, and early word learning. In M. Bowerman & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 215-256). Cambridge: Cambridge University Press.
- Fee, E. J. (1991). Underspecification, parameters, and the acquisition of vowels. Unpublished doctoral dissertation: University of British Columbia.
- Ingram, D. (1981). *Procedures for the analysis of children's language*. Baltimore, MD: University Park Press.
- Ingram, D. (1989). *First language acquisition: Method, description and explanation*. Cambridge University Press.
- Levelt, C. (1994). On the acquisition of Place. Doctoral dissertation. HIL Dissertations in Linguistics. Dordrecht: ICP Printing.
- Pye, C. 2020. Documenting the acquisition of indigenous languages. *Journal of Child Language* 1-26. [https://doi.org/10.1017/S0305000920000318].
- Pye, C., Berthiaume, S., & Pfeiler, B. (2020). Northern Pame-Spanish language acquisition in the context of incipient language loss. *International Journal of Bilingualism*, 24(2), 246–265.

Perl Programs

Concord.pl

#!/usr/local/bin/perl

#

concord.pl

Clifton Pye

This program produces a lexical concordance for a unicode text file exported from Elan via# Excel.

The exported text file should have a header row and be saved in the UTF-8 format with the .txt # extension

The header row should be: "part time child adult translation"

The program does not process blank lines

% words = ();

\$extension = ".txt";

\$concord = "con";

print "What is the file you wish to analyze? \n";

print "Type the filename WITHOUT THE EXTENSION .TXT \n";

print "The extension should be txt $n^{;}$;

<pre>\$textfile = <>;</pre>	# Read the filename from the keyboard input								
chop \$textfile;	# Remove line return								
\$file = \$textfile . \$extension;									
print "Now analyzing \$textfile\n\n";									
open text_in, "< \$file" or die("Cou	ıld not open \$file");								
<pre>\$line = <text_in>;</text_in></pre>	# Read header row								
<pre>chomp(\$line);</pre>	# Remove line returns								
<pre>@tiers = split ("\t", \$line);</pre>	# Count the number of columns								

```
open text_out, "> $textfile . $concord" or die( "Could not open the output file" );
if (scalar(@tiers) \geq 5) { #Process files with 5 column
 print text_out "Adult Word Child Word
                                                                          child utterances
                                            category
                                                            part
                                                                   time
                              translation\n"; # Print header
       adult utterances
 while (\  = < text in > ) {
   line_no = line_no + 1;
                                       # Count the line number
   chomp($line);
                                       # Remove line returns
   @tiers = split ("\t", $line );
                                         # Split line into tiers
        $child = $tiers[ 2 ];
         child = lc(child);
                                                # Change to lower case
        adult =  [ 3 ];
         adult = lc(adult);
                                                # Change to lower case
   $adult = ' ' . $adult . ' ';
                                          # Add spaces
   schild = \frac{s}{[''(), :.!?]}/g;
                                        # Remove punctuation
   adult = \frac{s}{[''(), :.!?]}/g;
                                        # Remove punctuation
   @chwords = split (" ", $child );
                                                  # Put words into an array
   @revchild = reverse( @chwords );
   @adwords = split (" ", $adult );
                                                  # Put words into an array
     foreach $adword (@adwords) {
        $chword = pop( @revchild );
        $words{$adword} = $words{$adword} . $chword . '
                                                                           '. $tiers[0].''.
$tiers[1].' '. $tiers[2].''. $tiers[3].''. $tiers[4].'#';
                                                                      # Put lines in hash
     } #end foreach word
 } #end while
} #end if
else { #Process files with 4 columns
```

part

time

adult translation";

Print header

print text_out "Word category

```
while ($line = <text_in>) {
  line no = line no + 1;
                                    # Count the line number
  chomp($line);
                                    # Remove line returns
   @tiers = split ("\t", $line );
                                      # Split line into tiers
       adult =  [ 1 ];
        adult = lc(adult);
                                            # Change to lower case
  $adult = ' ' . $adult . ' ';
                                       # Add spaces
  adult = \frac{s}{[''(), :.!?]} / g;
                                     # Remove punctuation
   @words = split (" ", $adult );
                                            # Put words into an array
    foreach $word (@words) {
       words \{ word \} = \
$tiers[3].'#';
                     # Put lines in hash
    } #end foreach word
 } #end while
} #end else
close text in;
# Display the results
foreach $word (sort keys %words) {
                                          # Sort the word hash
  @lines = split ("#", $words{$word});
                                          # Put lines into an array
 print text_out "$word";
                                             # Print word
   foreach $line (@lines) {
     print text_out " $line\n";
                                     # Print line
   } #end foreach line
} #end foreach word
```

Vowel.pl

#!/usr/local/bin/perl# vowel.pl# Created by Clifton Pye 2021

This program compares the vowels in a child's tokens with their adult targets# It analyzes data from a text file with two columns: adult child

\$extension = ".txt"; \$vowel = ".pon"; print "What is the file you wish to analyze? \n"; print "Type the filename WITHOUT THE EXTENSION .TXT \n"; print "The extension should be txt \n\n";

Read in the data file

```
$textfile = <>;  # Read the filename from the keyboard input
chop $textfile;  # Remove line return
$file = $textfile . $extension;
print "Now analyzing $textfile\n\n";
open text_in, "< $file" or die( "Could not open $file" );
$line = <text_in>;
open text_out, "> $textfile $vowel" or die( "Could not open the output file" );
# Read datafile
while ($line = <text_in>) {
    chop($line);
```

```
@words = split ("\t", $line );  # Split line into words
```

\$adult = \$words[0]; \$child = \$words[1]; \$child =~ s/[bdgptkmnlrwyzsfxch]+//g; # Delete consonants \$adult =~ s/[bdgptkmnlrwyzsfxch]+//g; # Delete consonants @vowel = split (" ", \$child); # Split vowels by spaces while (@vowel) { \$vow = pop(@vowel); if (\$vow != 7 && \$vow ne 'j') { \$vow =~ s/^7//; # Delete initial 7 \$vow =~ s/7\$//; # Delete final 7 \$vow =~ s/7\$//; # Delete final 7 \$vow =~ s/^j/; # Delete initial j

vow = s/j /; # Delete final j

 $\boldsymbol{vow} = \boldsymbol{vow} = \boldsymbol{vow} \cdot \boldsymbol$

}
}
@vowel = split (" ", \$adult); # Split vowels by spaces
while (@vowel) {
 \$vow = pop(@vowel);
 if (\$vow != 7 && \$vow ne 'j') {
 \$vow =~ s/^7//; # Delete initial 7
 \$vow =~ s/7\$//; # Delete final 7
 \$vow =~ s/7\$//; # Delete initial j

vow = s/j, # Delete final j

 $\$ sound hash

} } } close text_in;

```
# Display the results
print text_out "Child's vowels Adult word Child word\n";
@keys = keys( %childphon );
@sortkeys = sort @keys;
@reversed = reverse( @sortkeys );
while ( $key = pop( @reversed ) ) {
  @words1 = split( //, $childphon{$key} );
  @sorted1 = sort @words1;
  print text_out "\n$key";
  for ($i = 0; $i <= $#sorted1; $i = $i + 1 ) {
    print text_out " $sorted1[$i]\n";
  }
} #end child</pre>
```

print text_out "\nAdult's vowels Adult word Child word\n"; @keys = keys(%adultphon); @sortkeys = sort @keys; @reversed = reverse(@sortkeys); while (\$key = pop(@reversed)) { @words1 = split(//, \$adultphon{\$key}); @sorted1 = sort @words1; print text_out "\n\$key"; for (\$i = 0; \$i <= \$#sorted1; \$i = \$i + 1) { print text_out " \$sorted1[\$i]\n"; } } #end adult