



**A guide to preparing an Excel template
ready for upload to the OpenMWS platform**

Michela Santoro

With the assistance of Monica Lippolis for Part B

Under the supervision of the project co-ordinator

Prof. Anthony Baldry

Index

PART A: How to retrieve an mp3 file

PART B: Speech-to-text conversion to a Word table

**PART C: SIMPLIFICATION THROUGH MACROS: HOW TO SKIP A FEW
STEPS BY USING A MACRO**

PART D: TRANSFERING A WORD TABLE TO THE EXCEL TEMPLATE

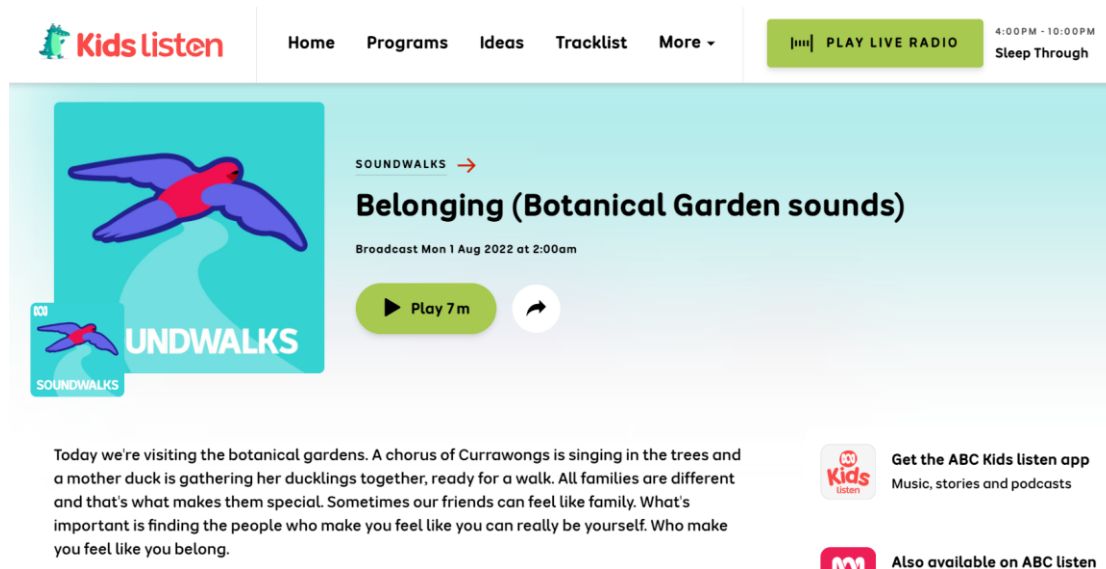
PART A: How to retrieve an mp3 file

Part A refers only to how to find the URL address in difficult cases almost always with MP3 audio-only files.

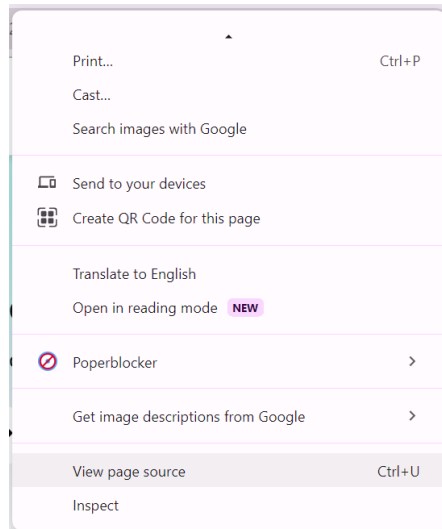
This guide explains how to retrieve an mp3 file from the podcast website and embed it in the OpenMWS Excel template's Overview "playlist" (Column I = Film Link).

For illustrative purposes ABC's Soundwalks <https://www.abc.net.au/kidslisten/programs/soundwalks/episodes> has been used.

STEP 1: Access the webpage of your podcast and open the page for the episode in question (in this case: <https://www.abc.net.au/kidslisten/programs/soundwalks/belonging/13987828>).



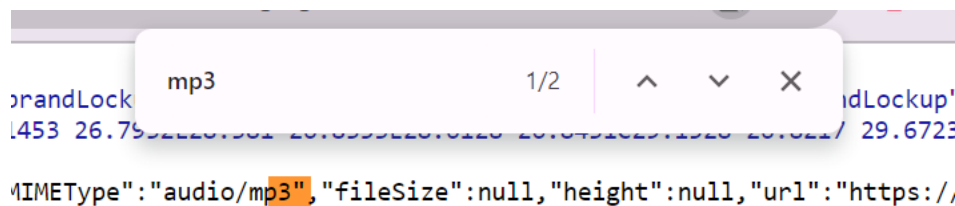
STEP 2: Right-click anywhere on the page and select "view page source".



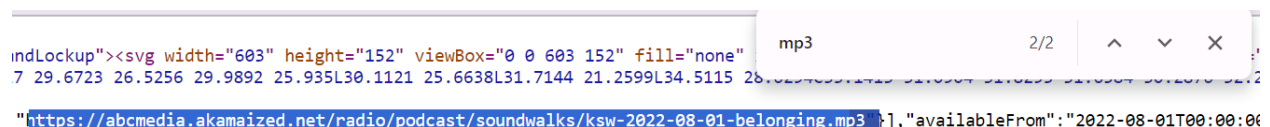
The following tab will open:

```
Line wrap
1 <!DOCTYPE html><html lang="en-AU"><head><meta charset="utf-8"/><link rel="preconnect" href="https://www.googletagmanager.com"/><link rel="preconnect" href="https://www.google
2 url(https://live-production.wcms.abc-cdn.net.au/42daa5095cd936669c5877e9eec078ea?impolicy=wcms_crop_resize&cropH=1125&cropW=1125&xPos=437&yPos=0&wid
3 First Australians and Traditional Custodians of the lands where we
4 live, learn and work.</p></div><div class="SiteFooterColumn_siteFooterColumn__RtVL" data-component="SiteFooterColumn"><h3 class="Typography_base__sj2RP Heading_heading__
```

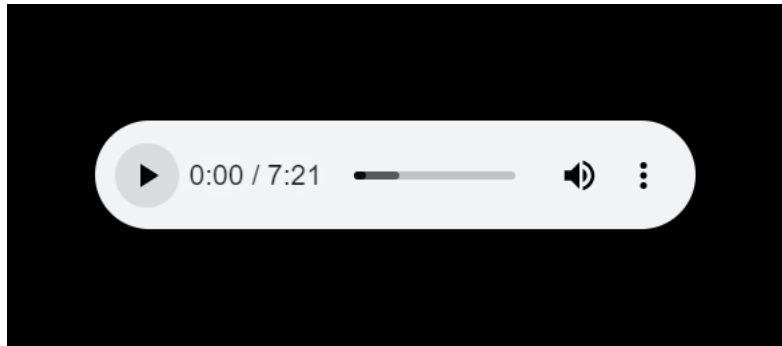
STEP 3: Search for mp3 using the browser searching tool (press “Ctrl + F” on your keyboard and type “mp3” onto the rectangle that will appear).



More than one result will often be highlighted. Choose the one where the mp3 marker is “preceded” by the *https* marker – in the current case, the second one.



STEP 4: Highlight the address, then copy and paste it onto a new browser tab to see if it works.



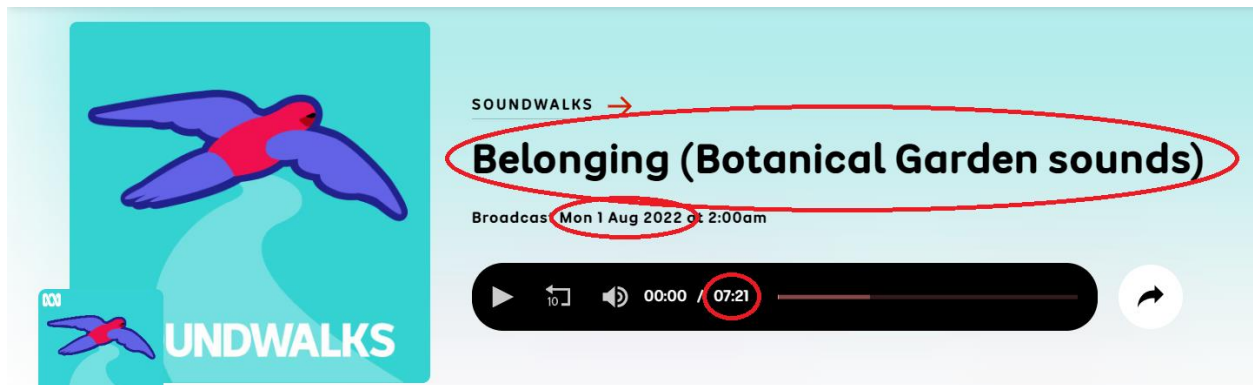
The screenshot shows the mp3 file thus obtained, downloadable by clicking on the three dots on the right and usable for any transcription.

STEP 5: Copy the address onto the OpenMWS Excel template.

The OpenMWS Excel template's Overview must be completed with the data for each podcast episode (see picture below).

VIDEO TITLE	FILM LINK	DURATION	DATE
Belonging (Botanical Garden sounds)	https://abcmmedia.akamaized.net/radio/podcast/soundwalks/ksw-2022-08-01-belonging.mp3	07:21	01.08.2022
Caring for others (Bush sounds)	https://abcmmedia.akamaized.net/radio/podcast/soundwalks/ksw-2022-08-01-caringothers.mp3	06:49	01.08.2022
Finding your voice (Promontory sounds)	https://abcmmedia.akamaized.net/radio/podcast/soundwalks/ksw-2022-08-01-findingyourvoice.mp3	08:08	01.08.2022
Teamwork (Island sounds)	https://abcmmedia.akamaized.net/radio/podcast/soundwalks/ksw-2022-08-01-teamwork.mp3	07:16	01.08.2022

All the information that you need can be found on the episode's webpage (see red circles in the picture below).

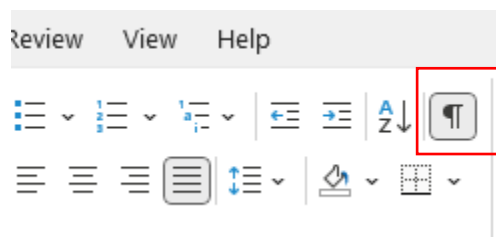


STEP 6: Copy the information that you need onto the Overview sheet. Alternatively, you can copy it onto a Word document and rearrange it on the OpenMWS Excel template.

PART B: Speech-to-text conversion to a Word table

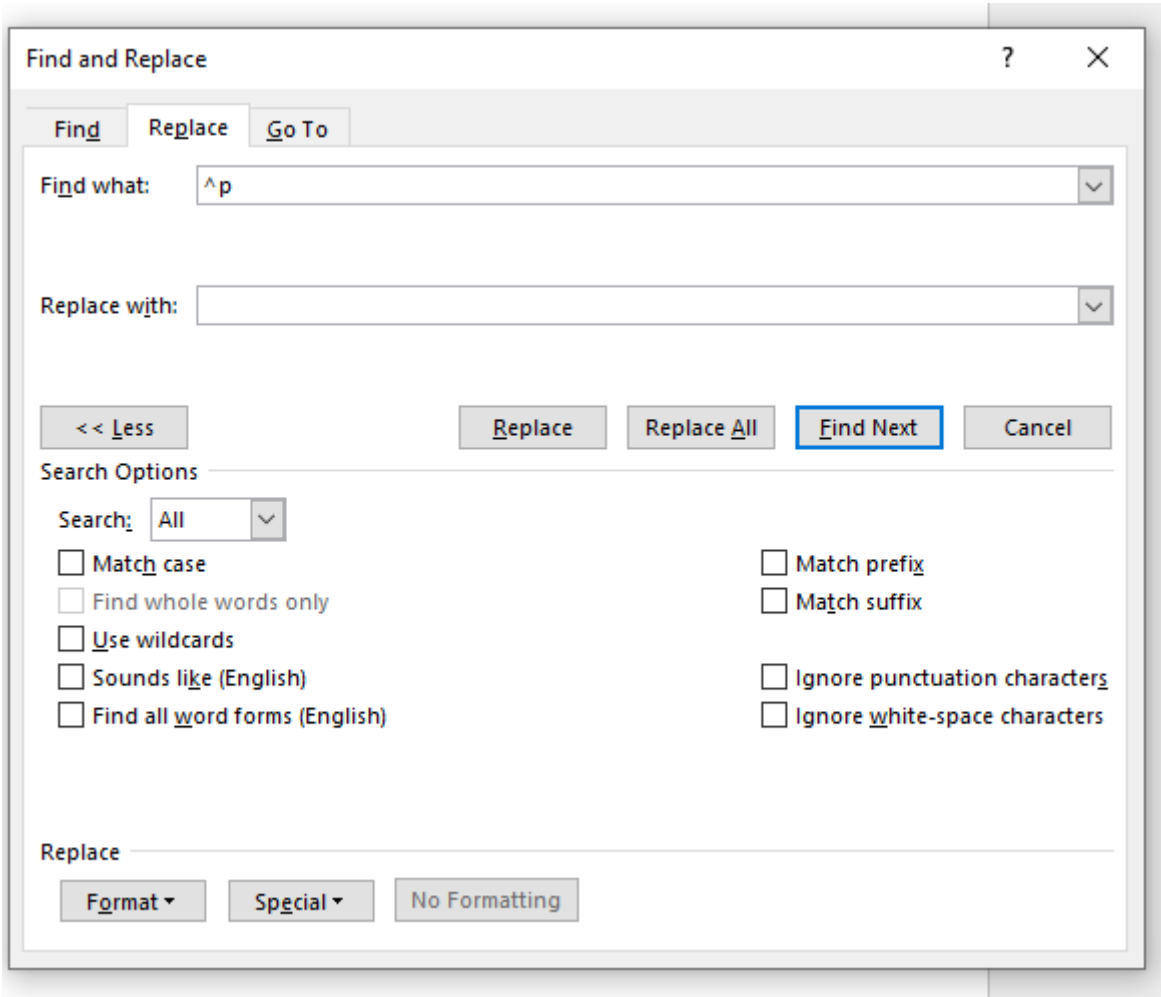
Background

- a) Familiarity with AI-based Speech-to-text conversion tools. See the manuals for speech-to-text conversion on the OpenMWS manuals page. When using TurboScribe adopt the following settings: US English or UK English; Speaker recognition: Detect automatically; when exporting choose the *with timestamps* option. If the transcripts are very short, group them together as an mp4 file using the free version of Bandicut (<https://www.bandicam.com/bandicut-video-cutter/>) and then uploaded to TurboScribe; Bandicut uses the mp4 format when joining files; use <https://www.freeconvert.com/mp3-to-mp4> for format conversion.
- b) **The problem to be solved:** manuscripts for plays and films are arranged as sets of speaker turns with a colon identifying the boundary between speaker name and speaker utterance in each turn; TurboScribe disrupts this convention by inserting timestamps between speaker identification and speaker utterance; this is out-of-step with the conventions used in plays and film scripts as well as those adopted in the OpenMWS project for speaker turns and turn duration;
- c) **The solution:** restoring the “traditional” conventions for speaker turns and turn duration facilitates overall online corpus management; one solution is to reorganise the data in the form of a two-column Word table: the first column is for time-stamps; the second is for the conventional structure of speaker turns: speaker identification + colon + speaker utterance;
- d) **The method:** is based on the assumption that punctuation is a metatextual system indicative of a transcript’s internal divisions and thus manipulates punctuation conventions in the TurboScribe transcript to achieve this reorganisation;
- e) **Method limitations:** the method is suitable for dialogues not monologues i.e. in cases where speaker *turns* are either absent or *limited*;
- f) **Further changes needed:** the method retains speaker identification as Speaker 1, Speaker 2 etc and does not convert speaker identification into a specific name; this is a separate step to be covered elsewhere;
- g) **Methodological preliminaries:** activate the paragraph symbol so that paragraph marks can be easily identified as you proceed.



STEP 1

Eliminate **ALL** paragraph marks in a TurboScribe transcript with timestamps, leaving **ONE** space (not shown in the screenshot) in the *Replace with* box. To do this go to Find and Replace and click Special. In *Find what* insert the first option Paragraph mark

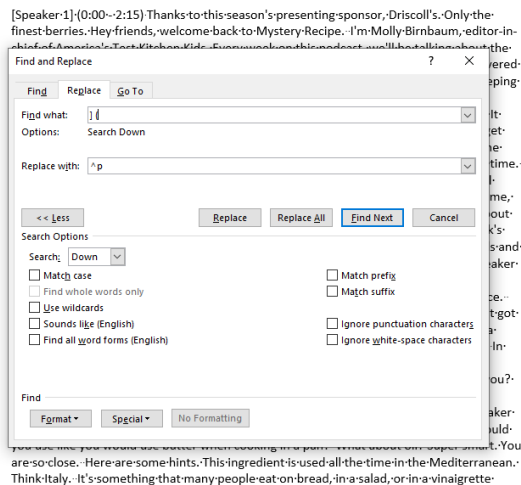


SHORTCUT METHOD FOR ADVANCED USERS ONLY

This shortcut eliminates the need to apply Steps 3 and 4; in particular it avoids the need to repeat speaker identification manipulation (Step 3) for every different speaker; however, it relies on the assumption that punctuation conventions remain constant; in other words, certain combinations such as **)**: i.e. *round bracket plus no space plus colon* are used exclusively to signal the *same* function throughout the text; this *may not* always be the case; when using this shortcut it is assumed that advanced users will be able to identify cases where this assumption is breached and to make appropriate adjustments.

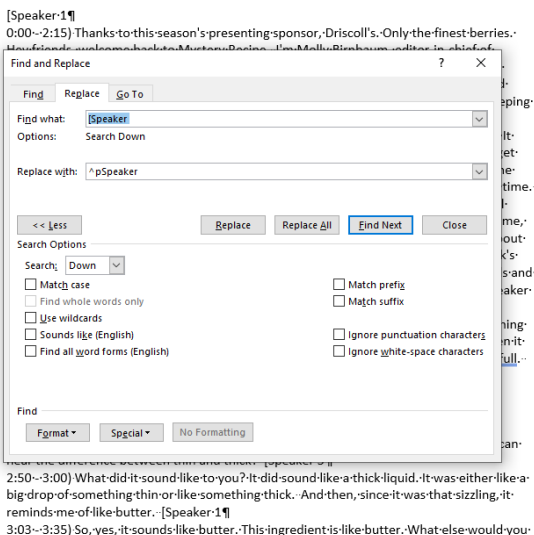
Shortcut Step 1

Use Word's find/replace function to replace **)** (with a **^p** in other words replace *closing square bracket + space + opening round bracket* with a *paragraph mark*.



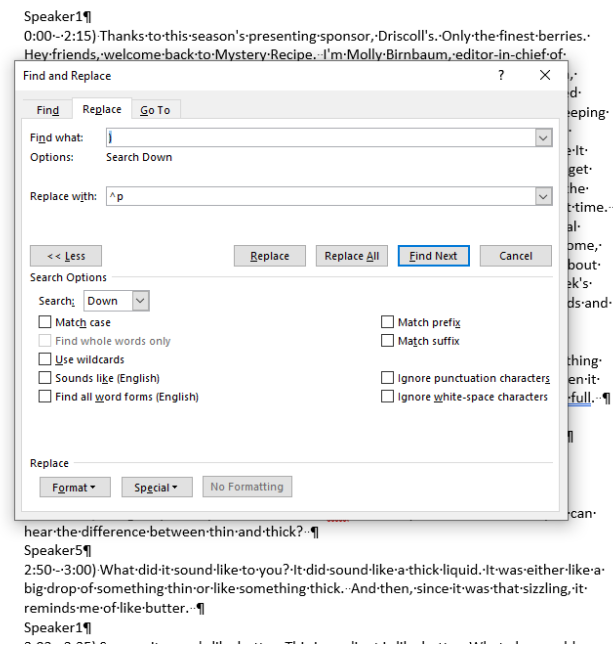
Shortcut Step 2

Replace **[Speaker** with a **^p Speaker** in other words replace *opening square bracket + Speaker* with a *paragraph mark + Speaker*; when doing so make the following adjustment delete the very first paragraph mark.



Shortcut Step 3

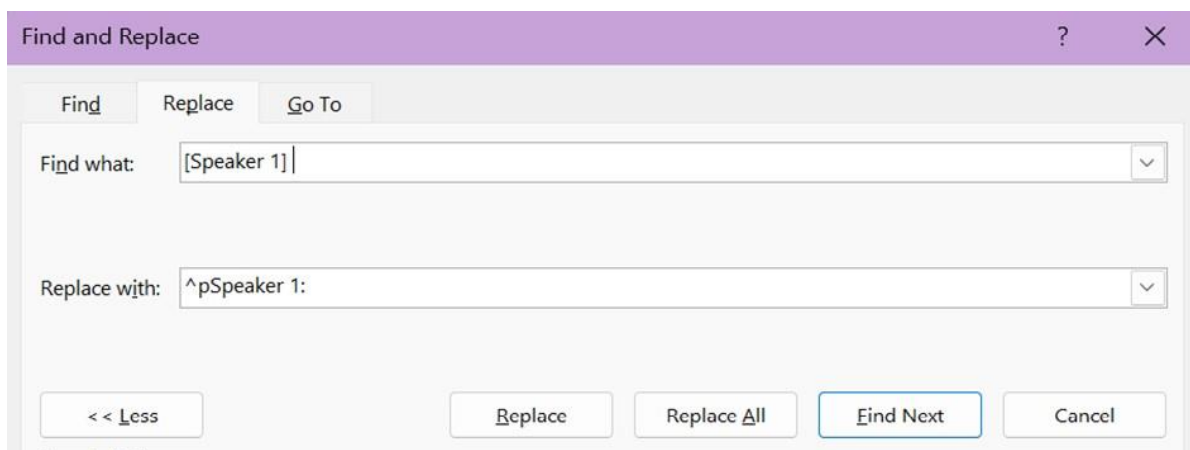
Replace `)` with a `^p` i.e. replace *a closing round bracket* with a *paragraph mark*.



Now skip Steps 2 and 3 and go to Step 4.

STEP 2

Using the same method, convert “[Speaker 1]” (do NOT forget spacing after the square bracket) to “^pSpeaker 1:” This creates a paragraph mark before Speaker 1.

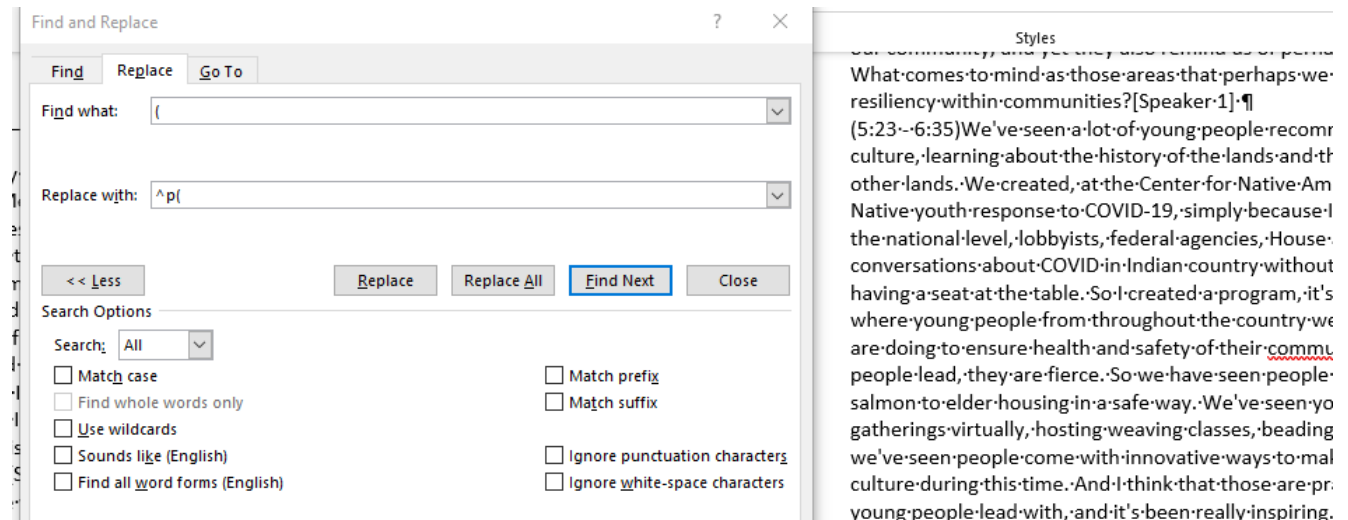


Apply this to all the speakers. This generates a paragraph mark before and after the speakers and will become the data for Column 1 for the table shown above in Step4.

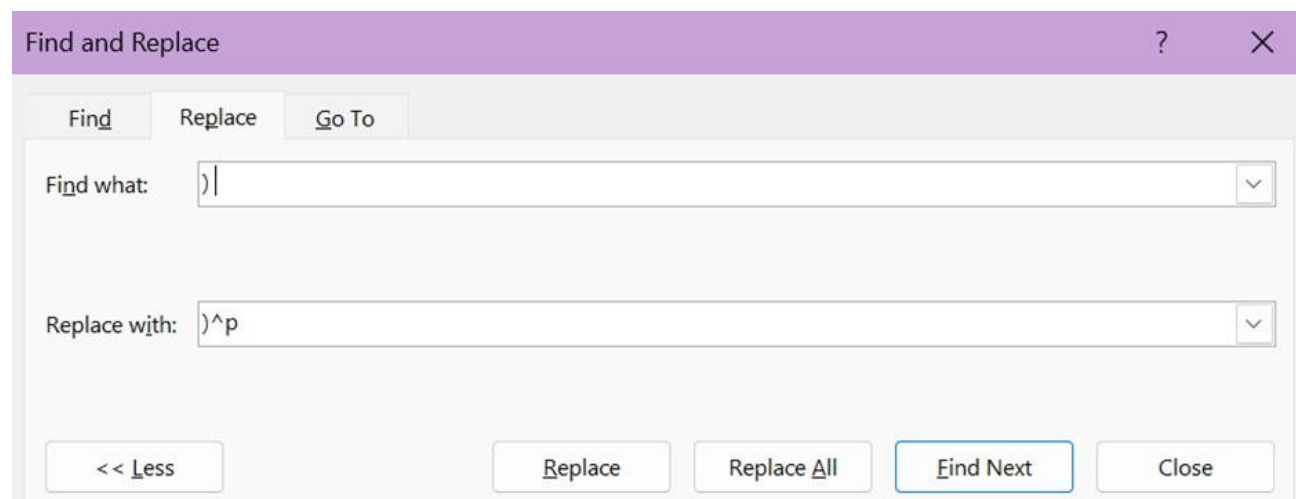
STEP 3

You now need to create paragraph mark that will become Column 2. This allows you to have the appropriate paragraph spaces. This is subdivided into two steps.

Step 3a: insert the bracket: “(” in *Find*, and “^p(” in *Replace with*

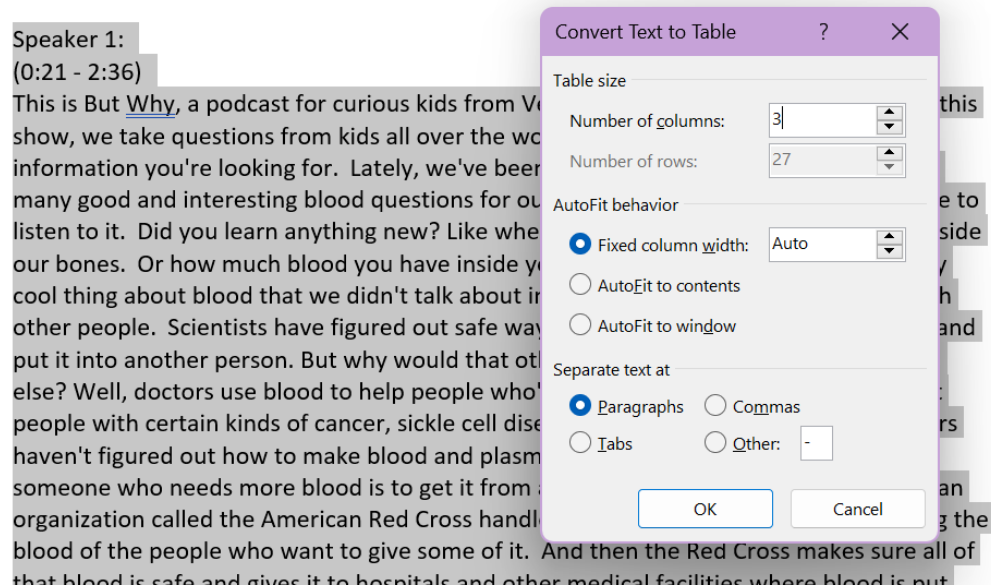


Step 3b: insert the bracket: “)” (do NOT forget spacing) in *Find*, and “)^p” in *Replace with*



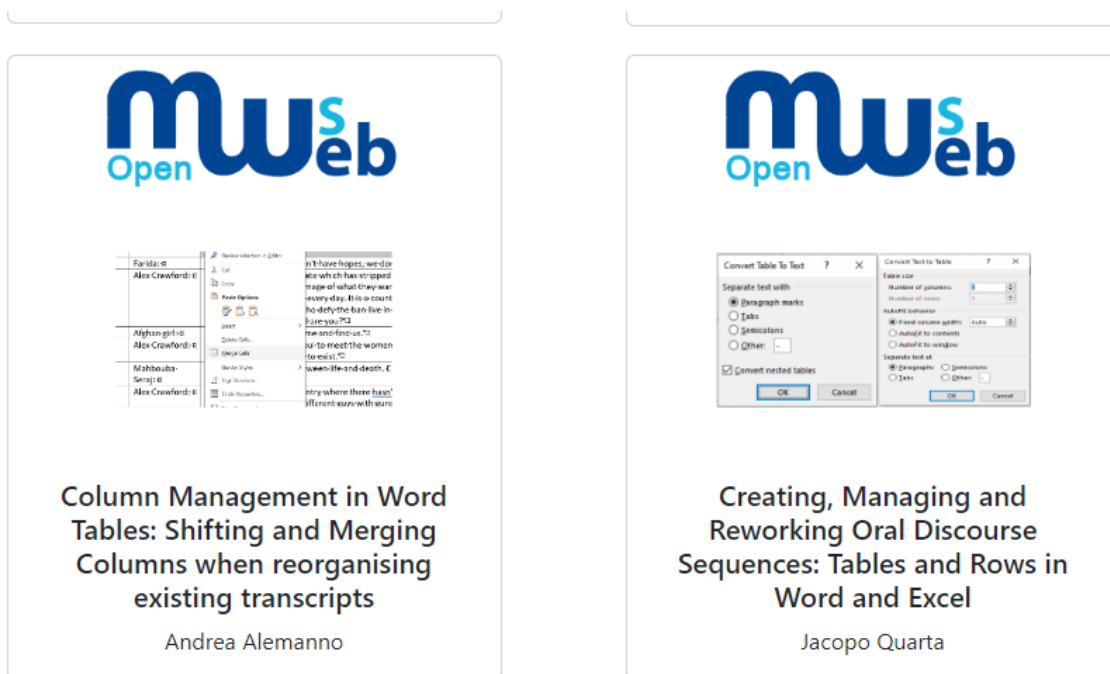
STEP 4

Select the entire text, click on “Insert”, then “Table” and use the Convert Text to Table function to create a Word Table. Note 3 columns.

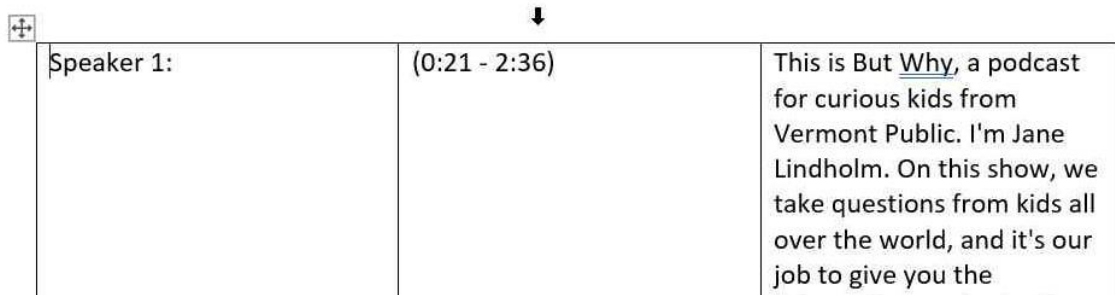


STEP 5

Use the BASIC manuals on the OpenMWS platform to shift the Time stamp column to the left. The steps to follow will be illustrated below.

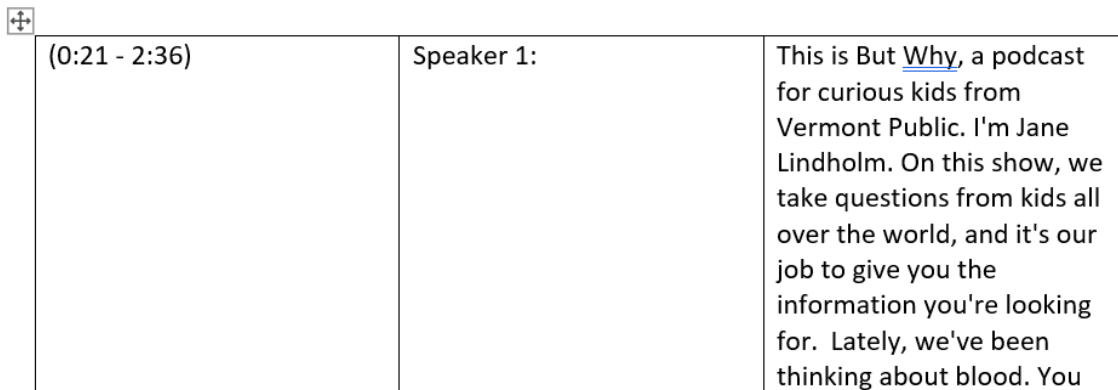


Step 5a: Place the cursor at the top of the second column until an arrow appears and left click it. This will select the text of the entire column.



Speaker 1:	(0:21 - 2:36)	This is But <u>Why</u> , a podcast for curious kids from Vermont Public. I'm Jane Lindholm. On this show, we take questions from kids all over the world, and it's our job to give you the
------------	---------------	--

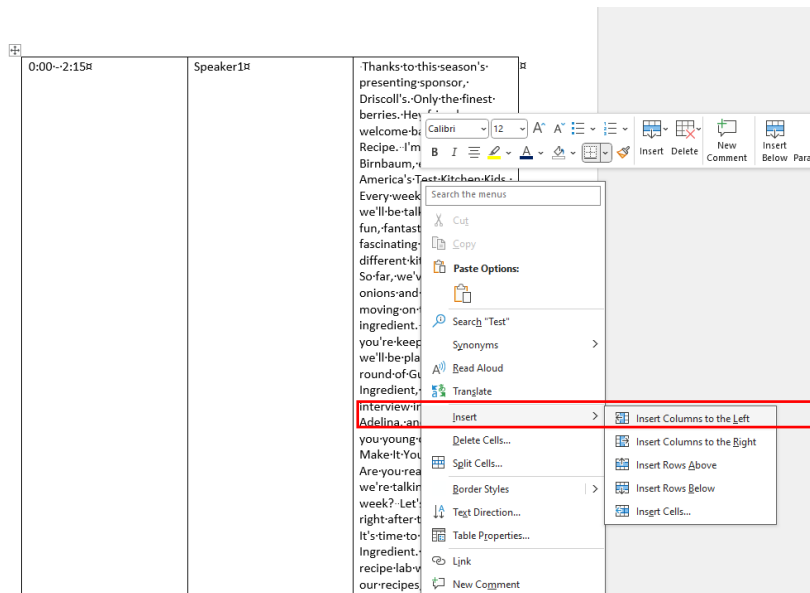
Step 5b: Once the text is selected, press and hold the left mouse button and shift it to the first column to invert the content of the two columns.



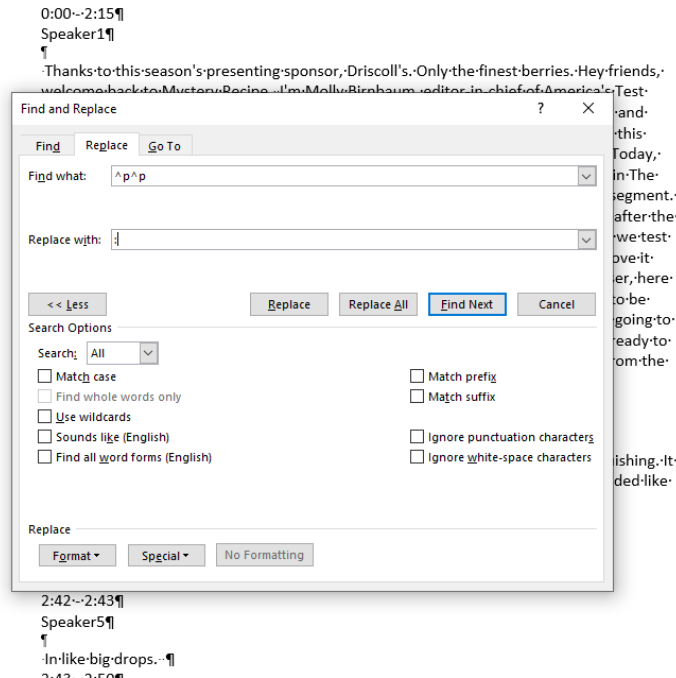
(0:21 - 2:36)	Speaker 1:	This is But <u>Why</u> , a podcast for curious kids from Vermont Public. I'm Jane Lindholm. On this show, we take questions from kids all over the world, and it's our job to give you the information you're looking for. Lately, we've been thinking about blood. You
---------------	------------	---

Second shortcut for advanced users

Create a new column between the Speaker Identification Column (Column 2) and Speaker utterance



This creates a blank column; then apply the Convert Table to Text functionality to reconvert the table to text. This produces two paragraph marks after speaker identification making it possible to apply the following change

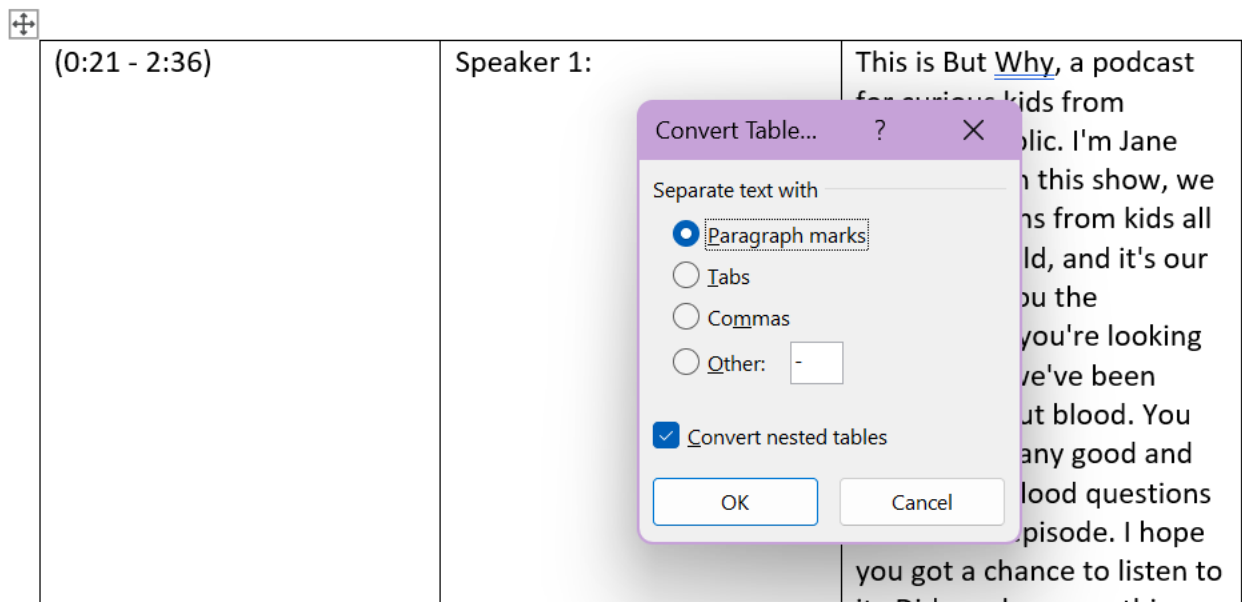


The two paragraph marks are replaced with a colon thus unifying speaker identification and speaker utterance in a single block without the need to apply separate changes for each speaker while keeping the time stamp block separate from the speaker identification/utterance block.

Now skip Steps 6 and 7 and go to Step 8.

STEP 6

Then click on “Layout” (you will see two; click on the one next to “Table Design”) and use the Convert Table to Text functionality to reconvert the table to text



Your text should look like this:

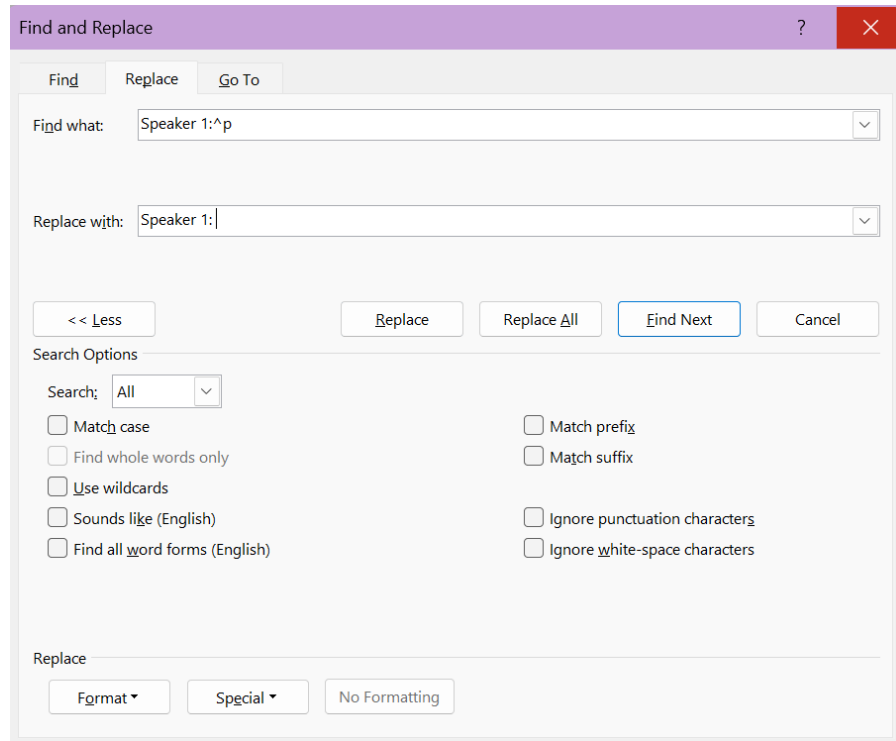
(0:21 - 2:36)

Speaker 1:

This is But Why, a podcast for curious kids from Vermont Public. I'm Jane Lindholm. On this show, we take questions from kids all over the world, and it's our job to give you the information you're looking for. Lately, we've been thinking about blood. You sent us so many good and interesting blood questions for our last episode. I hope you got a chance to listen to it. Did you learn anything new? Like where our blood is made? Spoiler alert, inside our bones. Or how much blood you have inside your body right now? There's one really cool thing about blood that we didn't talk about in the last episode. We can share it with other people. Scientists have figured out safe ways to take the blood from one person and put it into another person. But why would that other person need blood from someone else? Well, doctors use blood to help people who've been in accidents and to help treat people with certain kinds of cancer, sickle cell disease, and other medical issues. Doctors

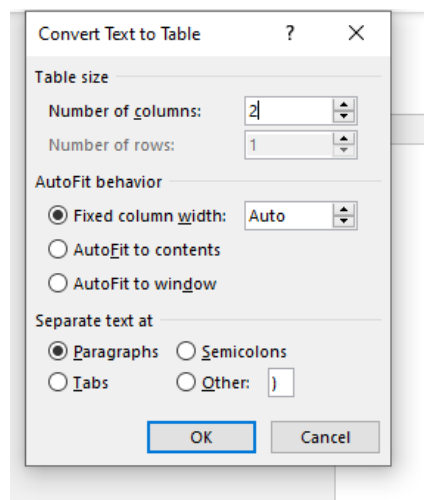
STEP 7

Now apply the following conversion to all the Speakers (the example shows just Speaker 1). Do NOT forget spacing in “Speaker 1: NOTE: it is



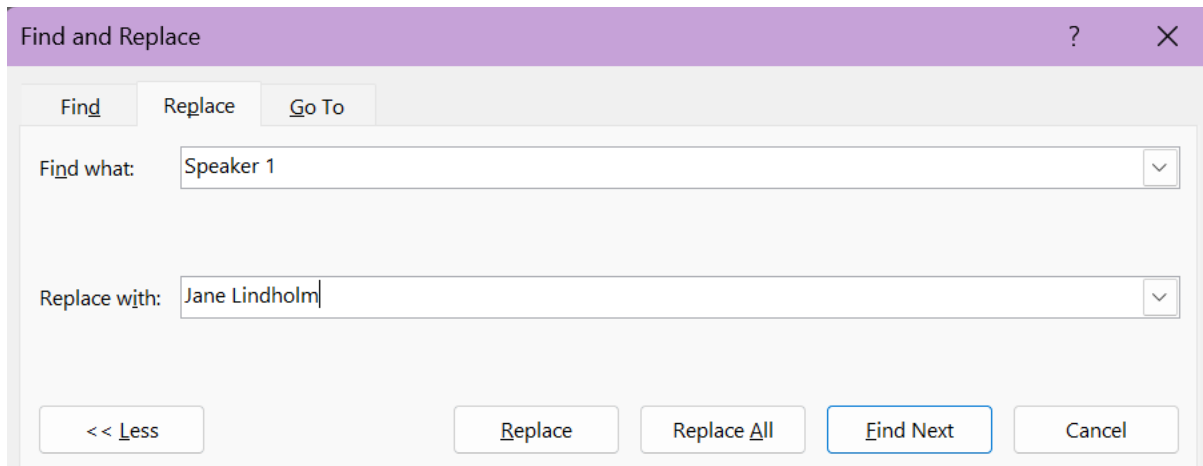
STEP 8

Now select your text and use the Convert Text to Table function to create a Word Table (see step 4). Note this time **2 columns**.



STEP 9

Using the same Find/Replace technique, change Speaker 1, Speaker 2 etc to their real names.



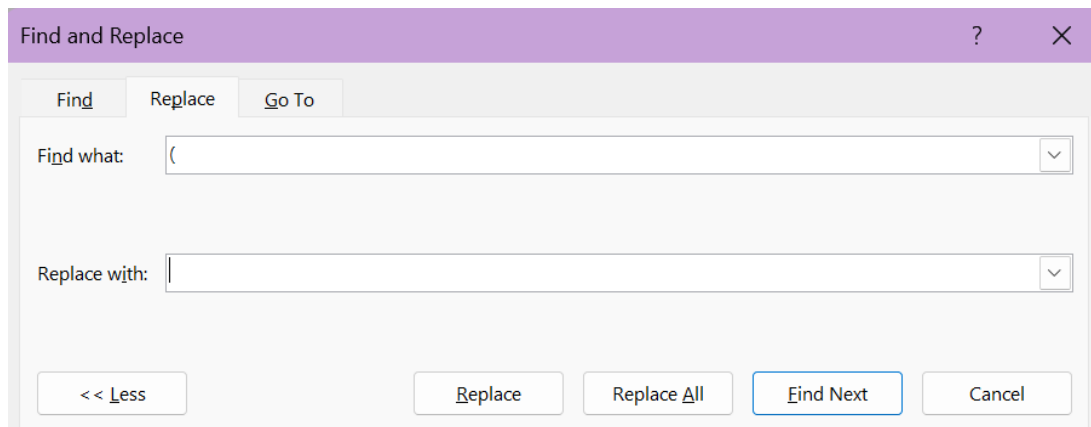
STEP 10

The second column can now be pasted onto the Excel file of the corpus you are working on, while the first column needs to be further adjusted.

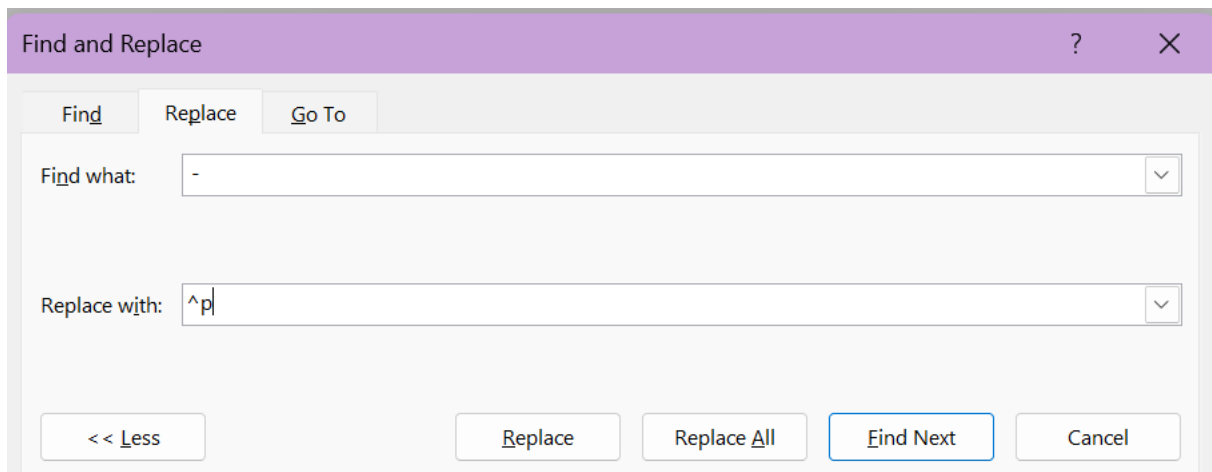
(0:21 - 2:36)	Jane Lindholm: This is But <u>Why</u> , a podcast for curious kids from Vermont Public. I'm Jane Lindholm. On this show, we take questions from kids all over the world, and it's our job to give you the information you're looking for. Lately, we've been thinking about blood. You sent us so many good and interesting blood questions for our last episode. I hope you got a chance to listen to it. Did you learn anything new?
---------------	--

STEP 11

Copy Column 1 into a separate file (it must NOT be in table form) and use the find/replace function to remove the brackets by typing “(“ and then “)” onto “Find what:”, leaving “Replace with:” blank both times.

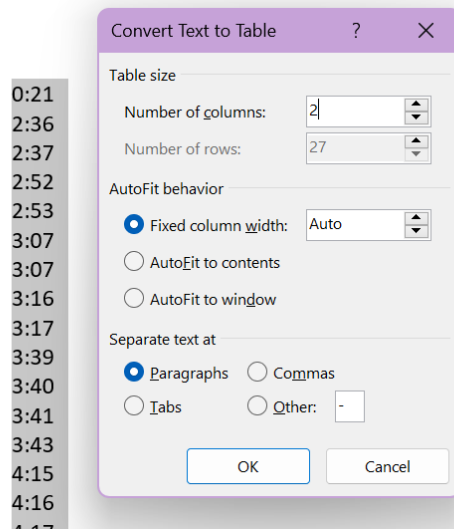


Then, replace the dash with a paragraph mark, by typing “ – ” (mind the spacing before and after the dash) and “^p” onto “Replace with:”.



STEP 12

Apply the *Convert Text to Table* function (number of columns: 2) and then delete the second column.



Adjust the first column so that the time stamp has the MM:SS structure (it should look like the table in the picture below). Paste the column in the original file. The conversion of a TurboScribe file to a two-column Word Table with accurate speaker names is now complete.

00:21
02:37
02:53
03:07
03:17
03:40
03:43
04:16
04:17
04:20
04:21
04:24
05:11
05:22

This process creates two separate Word Tables whose data gets copied and pasted manually onto an Excel file. However, if you want to create a single two-column Word Table, you can use the following steps.

STEP 1: Open your timestamps file (see picture above) and click on the space outside of the table corresponding to the first row.

00:21
02:37
02:53
03:07
03:17
03:40
03:43
04:16
04:17
04:20
04:21
04:24

STEP 2: Go over to the transcription table, select the entire text, left click on “Copy”.

The screenshot shows a context menu for a text selection. The menu is titled "Review selection in Editor" and includes the following options: "Cut", "Copy", "Paste Options:", "Insert", "Delete Table", "Distribute Rows Evenly", "Distribute Columns Evenly", "Border Styles", "AutoFit", "Text Direction...", "Insert Caption...", and "Table Properties". The "Copy" option is highlighted. The text selection is from a transcription table with the following content:

Jane Lindholm: This is Bu
for curious kids from Ver
Jane Lindholm. On this st
questions from kids all ov
it's our job to give you th
you're looking for. Lately
thinking about blood. You
good and interesting bloo
our last episode. I hope y
listen to it. Did you learn
Like where our blood is n
inside our bones. Or how
have inside your body rig
one really cool thing abo
didn't talk about in the la
share it with other peopl

STEP 3: Paste the copied text onto the timestamps table.

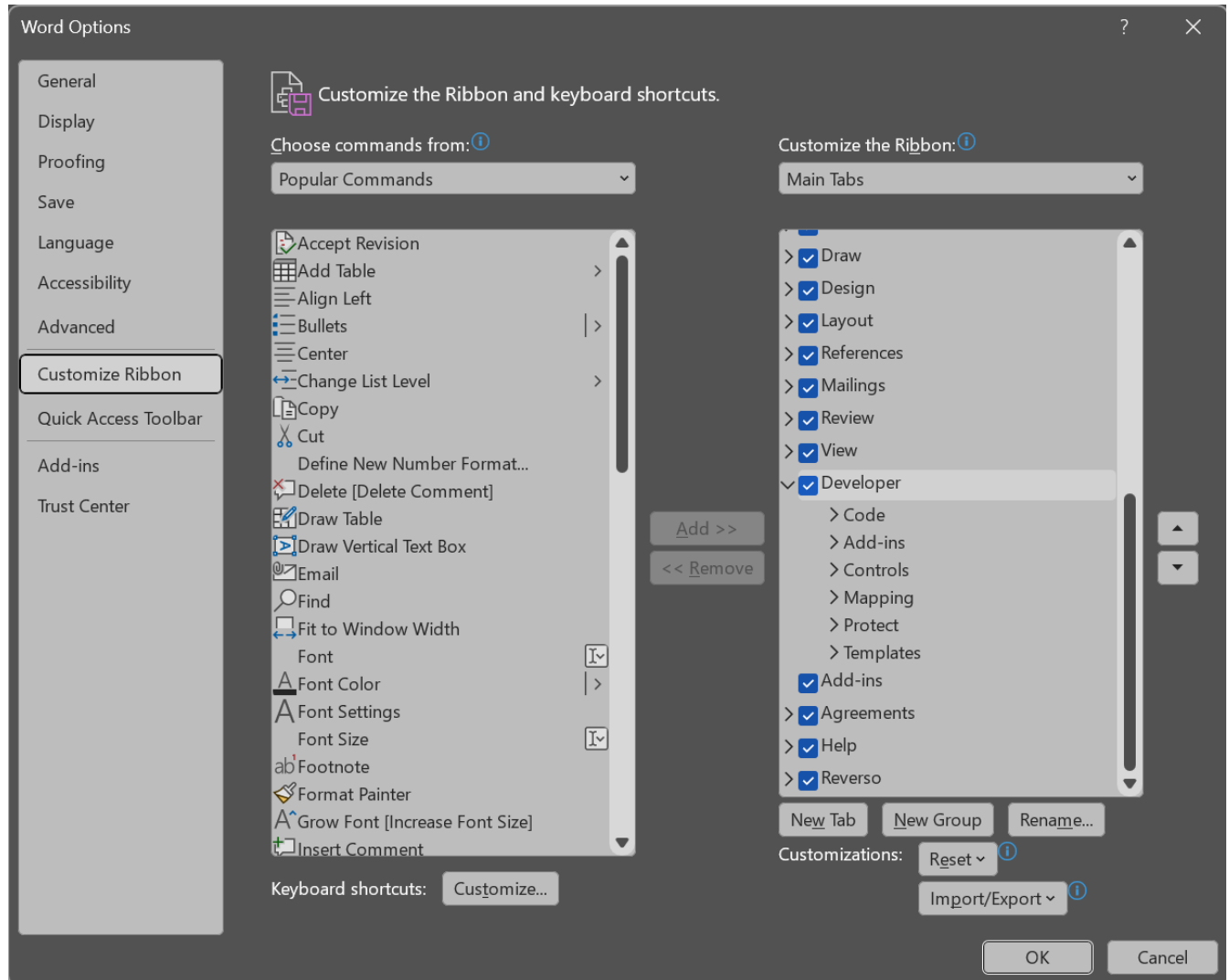
	Paste Options:  Paste (P)
00:21	Jane Lindholm: This is But Why, a podcast for curious kids from Vermont Public. I'm Jane Lindholm. On this show, we take questions from kids all over the world, and it's our job to give you the information you're looking for. Lately, we've been thinking about blood. You sent us so many good and interesting blood questions for our last episode. I hope you got a chance to listen to it. Did you learn anything new? Like where our blood is made? Spoiler alert, inside our bones. Or how much blood you have inside your body right now? There's one really cool thing about blood that we didn't talk about in the last episode. We can share it with other people. Scientists have

These steps will result in a two-column Word table with timestamps and their corresponding lines.

	do like a medical screening to make sure that she's eligible to donate.
02:53	Jane Lindholm: The questions were about my health, and they try to make sure that no one else is able to hear that conversation, so you can be completely honest in answering the questions. So I didn't record that part. Next up, they had to prick my finger to capture a little drop of my blood.
03:07	Linda Sweet: Testing the oxygen in your blood, so your hemoglobin, it's a different way of measuring, and it has to be 12.5 or higher for women to donate.
03:17	Jane Lindholm: Hemoglobin is a big word. That just means the amount of oxygen in your blood? Yes. It's a way of measuring it, the oxygen in your blood. It's like a little pinch, kind of. Already done. And you got a little slide with just a little tiny bit of my blood, and you're putting it in a machine, and that machine is going to tell you whether I have enough oxygen to donate today. Yes.
03:40	Linda Sweet: You have enough iron in your blood. Yep.
03:43	Jane Lindholm: All right. And it's very good.

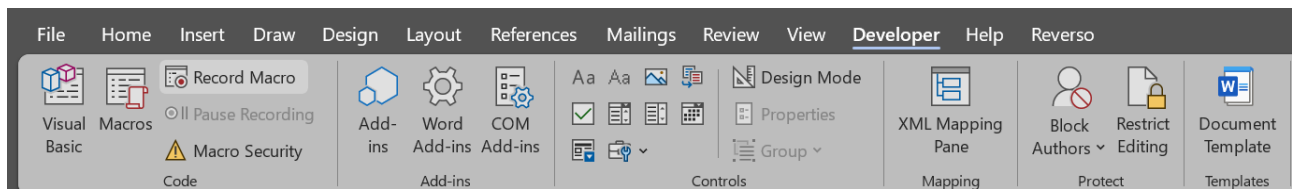
PART C: SIMPLIFICATION THROUGH MACROS: HOW TO SKIP A FEW STEPS BY USING A MACRO

The conversion of a TurboScribe transcript to a Word table can be a long process. Nevertheless, there are a few interesting tricks to make the job easier and smoother. It takes just a few moments to learn how. Word has a functionality called “macros” that allows users to record the actions used to edit a specific file that can then be applied to similar files requiring the same actions. This guide will show you how to do this, by breaking the process down into a few steps.

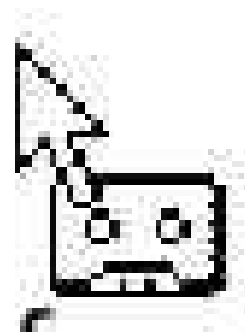
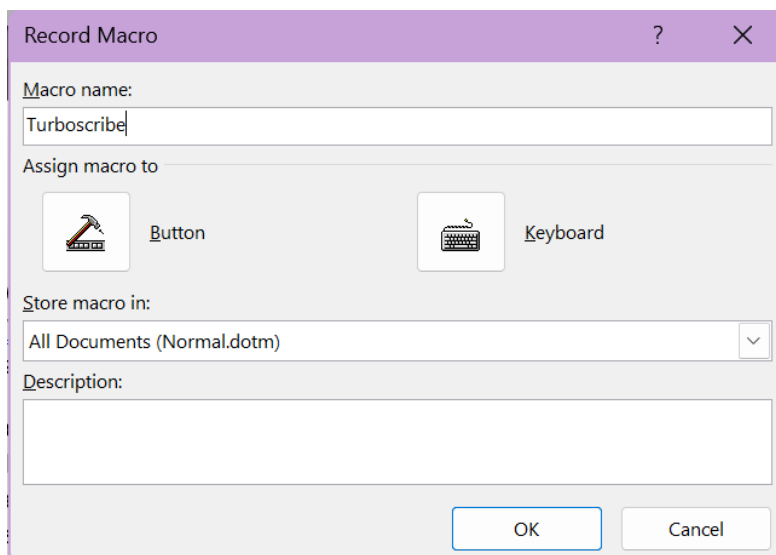


STEP 1: Open your Word transcript and click on the “Developer” tab at the top of your screen. If you do not have that, click on “File”: a list will appear. Then click on “Options”. This will open a dialog box. Now click “Customize Ribbon” and enable the “Developer” tab by ticking its checkbox (see picture below). This macro is now primed and ready to record your actions.

STEP 2: Select the “Developer” tab and click on “Record Macro”.

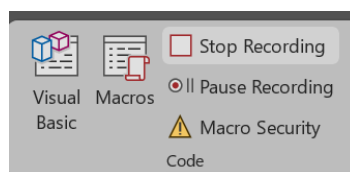


Choose a name for your macro such as “Turboscribe” (NOTE: spacing is NOT supported, so make sure you use a name that has no spaces, otherwise the program will label it as non-valid). Then press “OK” to continue. At this point you will see an icon (a cassette tape) next to your cursor, which means that your actions are being recorded (see picture on the right).



STEP 3: apply STEP 1, STEP 2 and STEP 3 (p. 2-3 above) to your document as you would normally have done had you been using the basic procedure.

STEP 4: once you are done, click on “Stop recording”.



Your macro is now recorded, and your text should look like this:

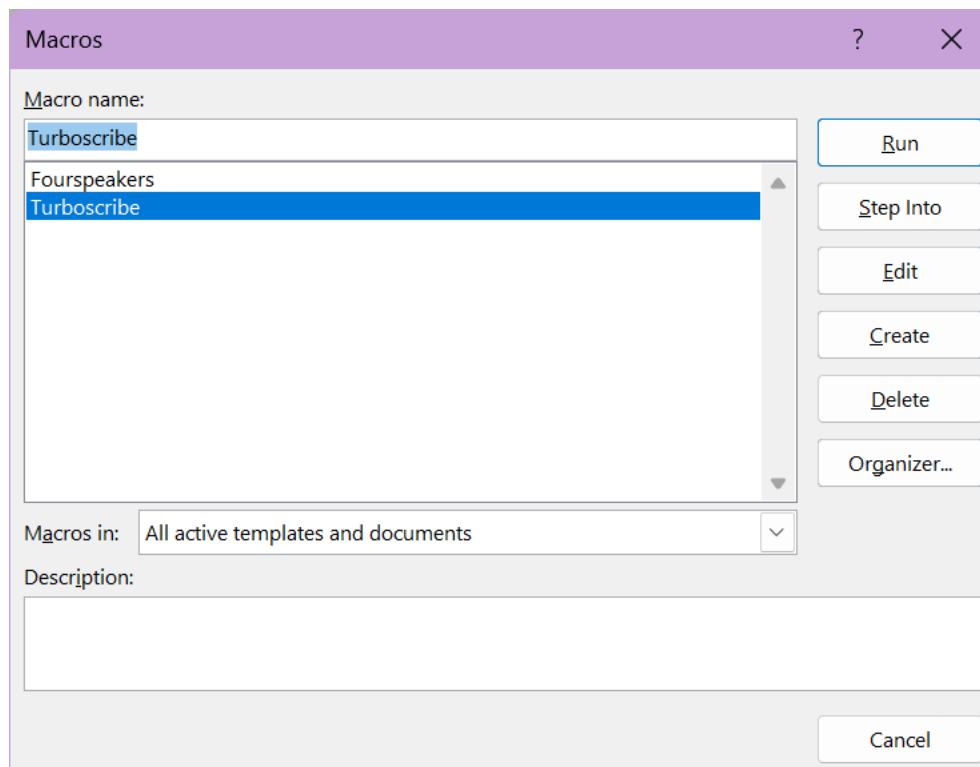
Speaker 1:
(2:12 - 2:22)
I run the Kidd Lab at UC Berkeley. So we specifically study infants and kids, but also adults because growing and learning continues throughout the lifetime.

Speaker 2:
(2:22 - 2:28)
Running the Kidd Lab sounds fun, but I'm picturing a big romper room with lots of toys. It's probably not what you do.

Speaker 1:
(2:28 - 2:50)
We do some of that. So one of the things that we study is how play leads to learning. So we do have a big romper room full of toys. We also make our own toys in order to study how kids deal with things that are new that they haven't seen before. How do you approach

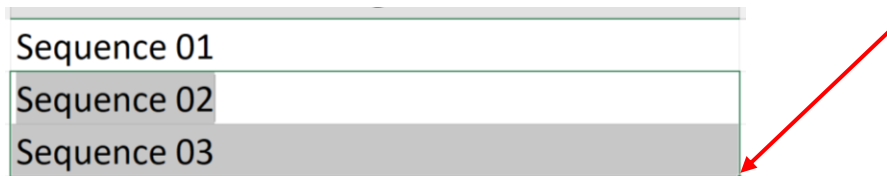
As this macro only simplifies part of the process, you must proceed manually with STEP 4 (p. 4) and the following steps until you get a two-column Word table (**STEP 4 – STEP 12** and **STEP 1 – STEP 3**).

However, now that this macro is recorded, whenever you need to modify a TurboScribe transcription, all you must do is access your document, click on the “Developer” tab, press “Macros”, select your labelled macro (in this case, “Turboscribe”) and press “Run”. Your document will be edited in just one click, and you will only have to add the remaining steps (**STEP 4 – STEP 12** and **STEP 1 – STEP 3**).

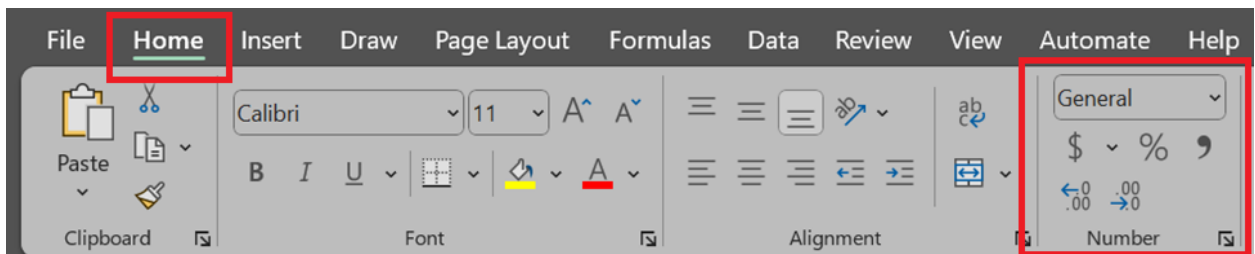


PART D: TRANSFERRING A WORD TABLE TO THE EXCEL TEMPLATE

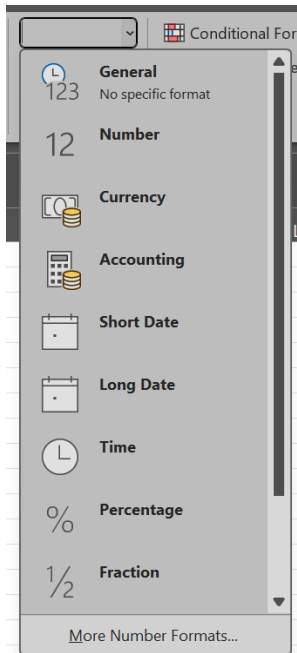
- 1- Paste the second column of the Word Table into Column F (Oral Discourse) in the second sheet of the Excel Template after the last completed row. OpenMWS automatically adjusts the order and exports the file with the correct ordering
- 2- Enter the item number in every row in Column A corresponding to each row pasted in. Tip: copy the item number from the OvSerview file
- 3- Enter Sequence 01 in Column B (or 001 in the case of an item with more than 99 rows). Use the pull-down technique to do this, i.e. placing the cursor on the bottom right-hand corner of the Excel cell and pulling down.



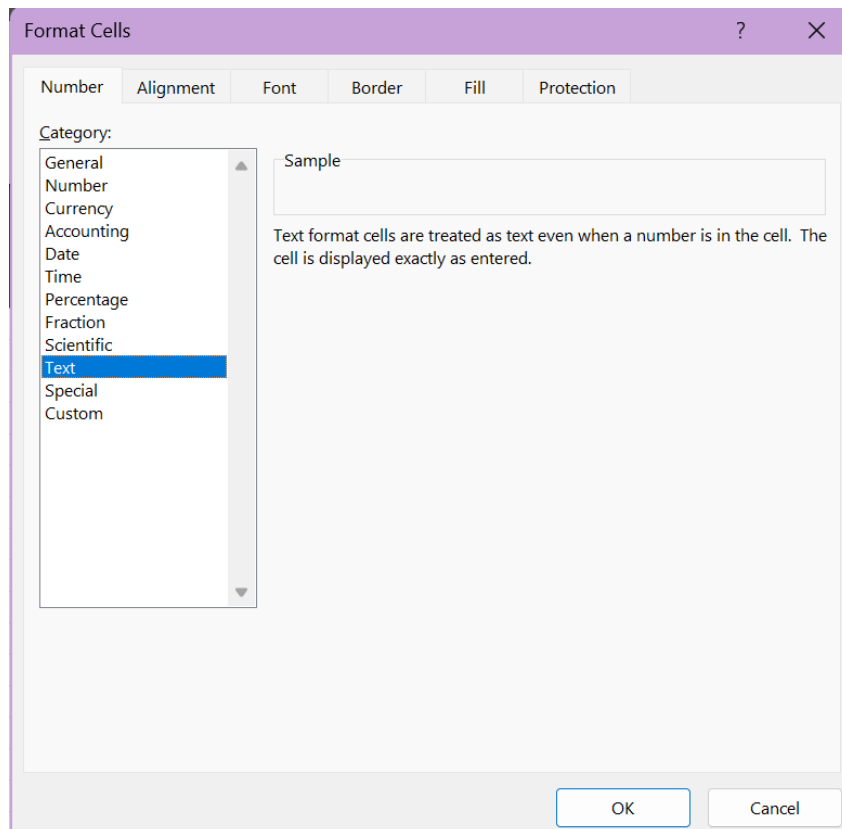
- 4- Format Column C so that it is in TEXT mode. Copy the timepoints in the Word Table into WordPad (for Windows platform or MAC and LINUX equivalents) and **de-format** the data before pasting into the Excel file
- 5- Alternatively, select Column C by clicking on “C”, then press “Home” at the top of your screen and find the “Number” box (see picture below)



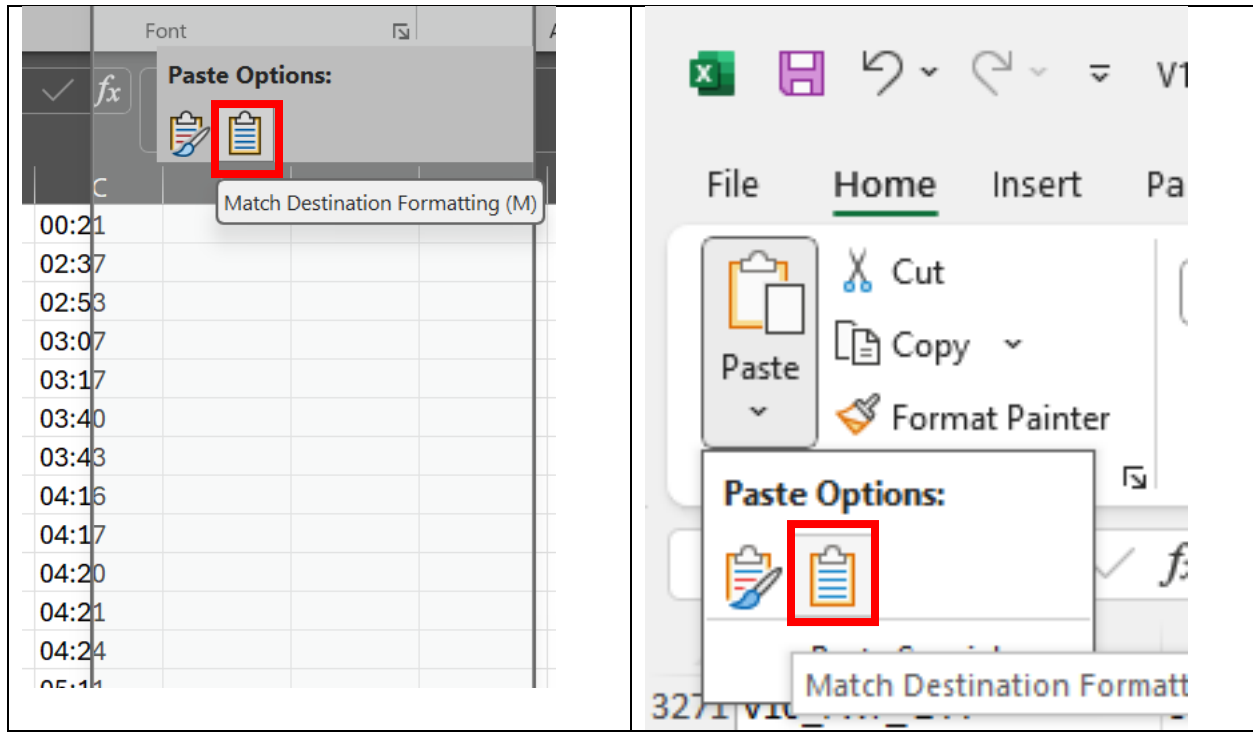
Click on the down arrow next to “General” to open the drop-down menu (shown) and select “More Number Formats...” to open a dialog box



Click on “Text” and “OK” to format Column C



- 6- Copy your timepoints (NOTE: they must appear as MM:SS). When pasting, right-click on the column C row corresponding to SEQUENCE 01 (or SEQUENCE 001) and click “Match Destination Formatting”. Timepoints will now be added correctly



- 7- Do not modify Columns D or E. Timespans will be automatically updated by OpenMWS on upload