MacroNarr v.14

https://macronarr.github.io/macronarr/index.html

A novel and easy way to analyse macrostructure in narratives

Documentation

Contents:

1. Summary of MacroNarr

Background

Research on narrative discourse has often focused on two levels of analysis:

- Microstructure: within-utterance or intra-sentential dimension (e.g. lexicon and syntax)
- Macrostructure: between-utterance dimension concerned with the integration of narrative discourse units (e.g. coherence)

At both levels, analyses of errors have often been employed (e.g. lexical errors, syntactic errors, coherence errors).

MacroNarr

Moving away from error analyses, **MacroNarr takes a novel frequentist approach** focusing on the level of macrostructure. Instead of deciding whether a part of the narrative contains an error, MacroNarr determines how “typical” each chunk of a narrative is in relation to reference data (e.g. baseline group).

2. Purpose and Use

The primary purpose of MacroNarr is to provide researchers with a novel, automated tool for analyzing macrostructural aspects of narrative discourse. By moving beyond traditional error analyses, MacroNarr adopts a frequentist approach to assess the typicality of narrative segments relative to reference data. MacroNarr can be used in research with various populations and across contexts. It offers an efficient solution to the quantification of macrostructural aspects of discourse (e.g. informativeness, coherence); these have often been measured inconsistently in previous research using variable, non-comparable analysis methods as well as scoring procedures (e.g. manual scoring which is time-consuming and subject to low inter-rater reliability).

Additionally, MacroNarr is under further development with the aim to have practical applications in clinical settings. The aim of this automated tool is to help clinicians streamline the scoring of narrative tasks (e.g. Cookie Theft) that are often used as part of neuropsychological assessment batteries with clinical populations (e.g. aphasia, dementia, schizophrenia). To achieve this, the MacroNarr team has laid out the following development stages:

- **<u>Development stage 1</u>**: We aim to create a database containing normative narrative discourse output. In other words, we will aggregate previously published narrative discourse data produced by healthy populations of children and adults on various tasks (e.g. Cookie Theft, Cinderella story, Dinner Party cartoon strip etc). Additionally, we will include in the database ready-made files (e.g., in simple, non-proprietary csv format) that will capture the frequency and order with which successions of Main Concepts (otherwise called narrative propositions) are typically mentioned by healthy controls. These will then be fully integrated with MacroNarr, so that they can be used as a reference point and compared against data from groups of patients or individual patients (similar to the *Single-Case Methodology in Neuropsychology* developed by John R Crawford; see here: https://homepages.abdn.ac.uk/j.crawford/pages/dept/SingleCaseMethodology.htm)

- **<u>Development stage 2</u>**: Following Stage 1, our goal in Stage 2 is to fully automate the MacroNarr analysis process, thereby enhancing accessibility and efficiency. We envision a system where researchers and clinicians can upload either an audio recording of narrative speech or a text transcription of the narrative, significantly reducing the manual workload currently required for analysis. To achieve this, we aim to rely on AI-based solutions capable of processing diverse forms of narrative input, such as natural language processing (NLP) and automatic speech recognition (ASR). From the text output (whether generated from ASR or directly uploaded), the system will employ NLP techniques to identify and extract narrative propositions. This can involve applying semantic similarity models, such as transformer-based models (e.g. BERT) fine-tuned on narrative datasets, to match extracted propositions against normative data. This process will also involve automatically organizing the extracted propositions and marking their order of mention. The MacroNarr tool can then compute the frequency and sequential order of propositions in the narrative input uploaded and processed on the website, comparing these to reference data from the normative database developed in Stage 1.

This dual focus on research and clinical utility highlights MacroNarr’s potential to bridge theoretical insights and real-world applications in the study and use of narrative discourse.

3. Analysis Basics

MacroNarr: Main Features

Based on previous research: MacroNarr uses the Main Concept (MC) analysis by Nicholas and Brookshire (1995). MCs are utterances with a single main verb, frequently produced by healthy individuals to communicate key elements of a stimulus (e.g. a picture). The aim of a MC analysis is to identify and evaluate utterances in discourse samples that match prespecified MCs, allowing researchers to establish to what extent core elements of a stimulus were communicated.

Taking a novel approach: MCs have typically been manually scored as absent, accurate/inaccurate, complete/incomplete etc. MacroNarr differs from previous approaches in two main ways:

(1) it quantifies the frequency of MCs (otherwise referred to as *narrative propositions*) mentioned in succession to determine how typical each chunk of a narrative is in relation to reference data, using concepts inspired by the Artificial Language Learning literature (Knowlton and Squire, 1994)

(2) it uses an automated analysis that requires minimal manual annotation by researchers.

MacroNarr: Steps and Key Terms

Below is a simplified outline of the steps taken to complete the analysis using the MacroNarr tool. While step 1 and 2 are to be completed by the researcher using this tool, Step 3, 4 and 5 are automated.

1. For the MacroNarr analysis, a list of MCs commonly mentioned by controls is needed. For some narratives, publicly available lists exists. For the Cookie Theft picture, see Nicholas and Brookshire (1995). For the Broken Window, the Cinderella, and the Peanut Butter and Jelly narrative tasks, see Richardson and Dalton (2016).

2. Once the list of MCs has been selected, the next step is to record the order with which single MCs (alternatively called *unigrams*) are mentioned by each participant in one's own research groups (target/experimental and a reference/control group).

3. Afterwards, the next step is to record the frequency with which combinations of two successive unigrams (called *bigrams*) and of three (called *trigrams*) were attested across the whole of the reference/control group.

4. Then, after we subtract -1 from each one of these frequency counts **(*)**, they become the values we assign to the same combinations that also appeared in each participant in the reference/control group and in the target/experimental group.

5. By averaging across bigram and trigram values recorded for each participant, we obtain individuals' Associative Chunk Strength (ACS) scores (Knowlton and Squire, 1994), a measure of typicality (in terms of narrative structure in this case).

(*) Subtracting -1 from each frequency count corrects for the fact that every combination produced by a control is counted at least once, ensuring fairer comparisons with participants in the target/experimental group by removing this baseline bias.

4. Instructions for Files

Files and format

Your files for the MacroNarr analyses must be in simple csv format. No other file types are allowed for upload. The two required files are the following:

(1) a file containing the order of narrative propositions mentioned by each participant in a target/experimental group.

(2) a file containing the order of narrative propositions mentioned by each participant in a reference/control group.

Major Considerations

In these files, **the first column should always be made up of unique codes** that are (or include) string-based input, i.e., text. **These represent individual narrative propositions** (e.g. prop_a, prop_b, MC1, MC2 etc). If you do not include string for labelling unique propositions (e.g. if you use numbers like 1, 2, 3 throughout the column), then in the output generated by MacroNarr these number-based proposition codes will be converted to the format 'mc_(n of row)'. For example, the new code “mc_4” will be assigned to the original proposition code that was found in the 5th row (i.e. 4th observation, excluding header) of the first column in your file. You will also get a warning if this occurs, as shown below in Figure 1.

Figure 1

[File labelled “test_file_1.csv”]

Proposition_codes	fake_participant1	fake_participant2	fake_participant3
1			
2		1	1
3	1		2
4	2		
5			3
6	3	2	4

NOTE: In your file test_file_1.csv, the first column contained input other than text. Adding column with input per row as 'mc_(n of row)'.

Additionally, while the first column is automatically assumed to contain unique codes for narrative propositions, **all other columns are assumed to represent data from unique participants**. In each participant’s column, you should mark the order in which the narrative propositions were mentioned using numbers, in ascending order of mention. This means that a "1" indicates the first mentioned proposition, "2" indicates the second mentioned proposition, and so on. If a proposition is repeated within the narrative, you can mark the order by appending a "+" symbol, as in "1+5+9," where "1" indicates the first mention for that particular proposition, “5” the second mention (first repetition) and “9” the third mention (second repetition). See Figure 2 below for an example.

Figure 2

Proposition_codes	fake_participant1	fake_participant2	fake_participant3	
a				
b		1	1	
c	1		2	
d	2+4			
e			3	
f	3	2	4	

First-row column headings can be determined by you and do not need to have a standard label.

For marking the order of propositions in participants' columns, **ensure that only numbers are used**. The "+" symbol is also allowed, and is always assumed to be a marker of repetitions. If you include string-based input, i.e., text, these observations will be excluded from analyses and you will also get a warning. See Figure 3 below for an example.

Figure 3

[File labelled "test_file_2.csv"]

Proposition_codes	fake_participant1	fake_participant2	fake_participant3	
a				
b	f12	1	1	
c	1		2	
d	2+4	string		
e			3	
f	3	2	4	

NOTE: Ignoring the first column and any cells with '+' in your file test_file_2.csv, the following columns have values with text and these values have been excluded from calculations: {'fake_participant1': ['f12'], 'fake_participant2': ['string']}.

Furthermore, you need to ensure that the **same unique codes** for propositions are used in the first column of **both your target and reference file**. If there are differences in this respect, you will get a warning, as shown below.

Figure 4

[File labelled “test_file_3.csv”]

Proposition_codes	fake_participant1	fake_participant2	fake_participant3
l			
m		1	1
n	1		2
o	2		
p			3
q	3	2	4

NOTE: Your file test_file_2.csv and your file test_file_3.csv do not have the same code labels. ACS calculations may be wrong...

Moreover, **it is advised that you use informative names to label your two files**. If your files are identically named, this is not a problem for MacroNarr and analyses will proceed as normal. You will, however, get a warning to inform you of the identical labelling in case you accidentally uploaded duplicate files (see below).

NOTE: Your file test_file_3.csv and your file test_file_3.csv have the same name. Did you mean to upload different files? Proceeding with calculations...

Finally, if you forget to upload one or both files, analyses cannot proceed, and you will also get a warning.

One file is missing....Please upload the other file.

Both files are missing....Please upload the two files.

Other considerations

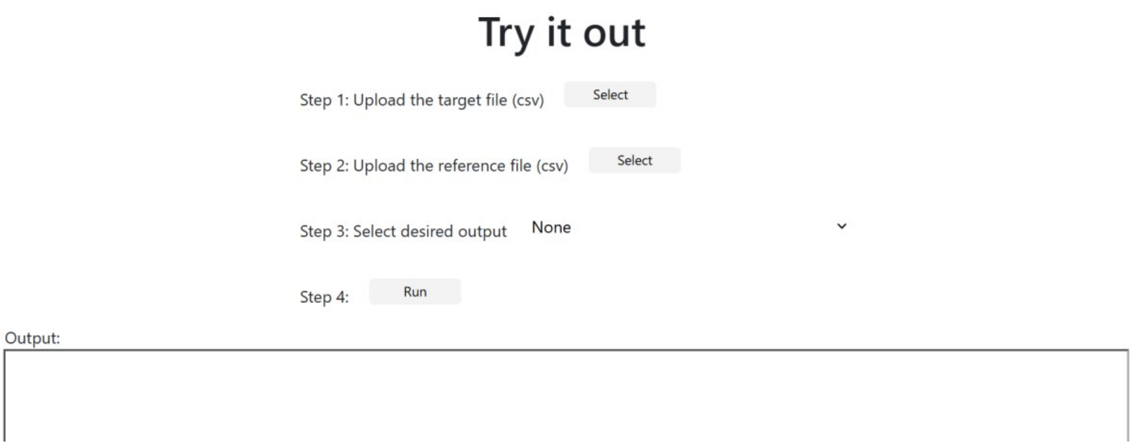
If there are missing values in the first column (unique codes for narrative propositions), MacroNarr interprets them as `None` and does not assign meaningful proposition codes. Users are advised that the **first column should not contain any unpopulated rows in between populated rows**, as this may affect the accuracy of the analysis.

If the target and reference files differ in the number of rows, MacroNarr will adjust the lengths to ensure compatibility for comparison. Users should know that this adjustment occurs automatically, and empty rows are padded with `None` or `n/a`.

5. “Trying it out” and Obtained Measures

To use the MacroNarr tool, click on the “Try it out” page of the MacroNarr website. Dummy datasets are available on the MacroNarr github repository. You will have to initially wait for ~20 seconds for the tool to load. This is because it is powered by Pyodide (CPython in WebAssembly), which loads a Python interpreter and supporting libraries directly into your web browser. This process involves downloading and initializing the necessary components in your browser, which can take some time depending on your internet speed and device capabilities. Once loaded, the tool will operate smoothly without requiring any downloads or installations. Figure 5 below shows the interface displayed after loading is completed.

Figure 5



Under Step 1, upload the file containing data from your target/experimental group.

Under Step 2, upload the file containing data from your reference/control group.

Under Step 3, select your desired output in the drop-down menu (see below for options).

Finally, **under Step 4**, click “Run” once you have completed the above steps.

MacroNarr analysis results will be displayed in the “Output” block-level container.

Obtained measures

The output options are:

- **All output**: All of the options described below.

- **Unigram count** (unique, total): Unigram count records frequency counts of single attested propositions and can be used as a **measure of informativeness** in narrative discourse production. It captures both the number of propositions attested once (unique count) as well as the total frequency of all attested propositions, including repeated ones (total count). This is a way to quantify the informativeness (i.e., richness, diversity of propositions) of narrative discourse samples. In parallel, it can identify less informative (e.g., fewer details of narrative content conveyed) or redundant narrative output through repetitions. See below for an example showing the two new columns that are created for each participant to represent the unique count (ending with "_Unique") and the total count (ending with "_Total").

REFERENCE (unigram count)

Participant	Count_Unigrams_Unique	Count_Unigrams_Total
fake_participant1	3	4
fake_participant2	2	2
fake_participant3	4	4

- **Bigrams and trigrams** (bg, tg): Bigrams represent pairs of propositions mentioned in succession (e.g. MC1-MC7, MC7-MC9 etc), while trigrams represent triplets of consecutive propositions (e.g. MC1-MC7-MC9, MC7-MC9-MC15 etc). This measure yields the identities of bigrams and trigrams mentioned by each participant, listed in the order in which they were mentioned. It has been included for the reference of the user, making it possible to inspect whether there are differences in the combinations of propositions that have been most frequently attested in different participant samples. See below for an example showing the two new columns that are created for each participant to represent bigrams (ending with "bg") and trigrams (ending with "tg").

REFERENCE (bigrams and trigrams)

Proposition_codes	fake_participant1bg	fake_participant1tg
a	cd	cdf
b	df	dfd
c	fd	
d		
e		
f		

- **Bg, tg & frequencies** (based on reference): This measure is related to the one above. The difference in this case is that bigram and trigram identities are listed not in the order in which they were mentioned by each participant, but based on frequency ranks as determined based on data in the reference/control group. To explicate, this measure is based on the frequency with which bigram and trigram identities were attested across the whole of the reference/control group. These frequency counts undergo a subtraction of -1 (see page 5 for justification); then, they become the values assigned to the same combinations that also appeared in each participant in the reference/control group and in the target/experimental group. Four columns are created for bigrams and trigrams for each participant: two which list bigrams/trigrams attested and sorted by frequency rank (ending with "bg/tg_sorted_by_frequency"), and two which show their corresponding frequency values based on data from the reference/control group (ending with "bg/tg_frequency"). See below for an example.

REFERENCE (bg, tg & frequencies, based on reference file)

fake_participant1bg_sorted_by_frequency	fake_participant1bg_frequency
df	3.0
fi	3.0
cd	2.0
wx	2.0
ij	1.0
sw	1.0
js	0.0

- **Associative Chunk Strength** (ACS): ACS is the final measure yielded based on the measures described above. It averages across bigram and trigram values assigned to each participant, and yields individuals' ACS scores (Knowlton and Squire, 1994), **a measure of typicality in terms of narrative organisation**. One column is created for ASC scores (labelled "Mean"). See below for an example.

TARGET (ACS)

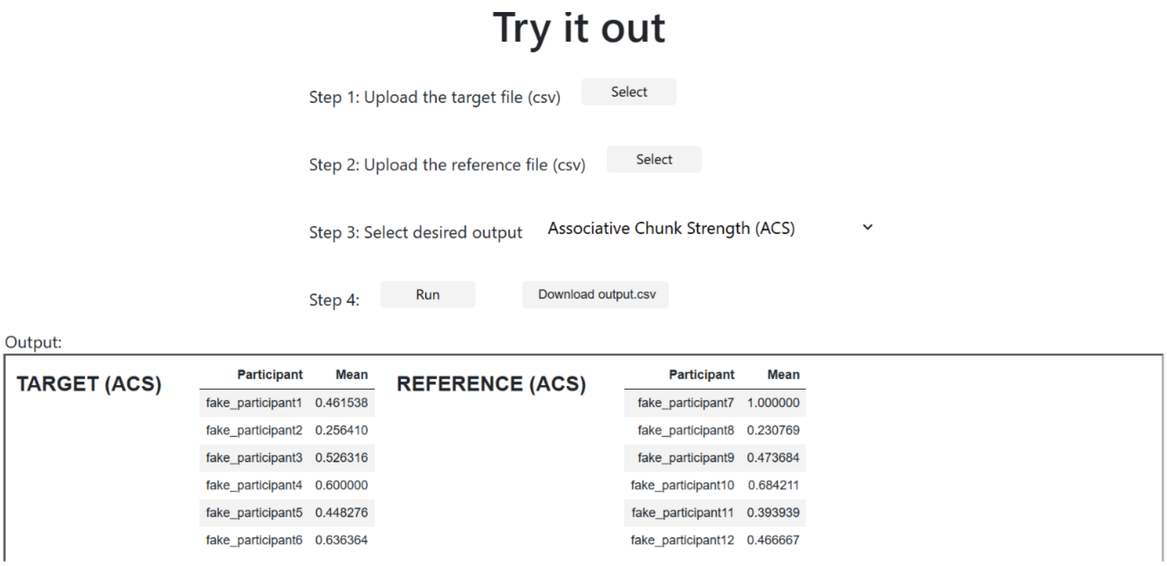
Participant	Mean
fake_participant1	0.461538
fake_participant2	0.256410
fake_participant3	0.526316
fake_participant4	0.600000
fake_participant5	0.448276
fake_participant6	0.636364

REFERENCE (ACS)

Participant	Mean
fake_participant7	1.000000
fake_participant8	0.230769
fake_participant9	0.473684
fake_participant10	0.684211
fake_participant11	0.393939
fake_participant12	0.466667

All the above measures are produced for both the reference/control group and the target/experimental group. They are displayed in the "Output" block-level container and can be downloaded by clicking the "Download output.csv" button as shown in Figure 6 below.

Figure 6



6. Implementation References

For implementation references, see:

Tsoukala, A., Hinzen, W., Rosemary, V., & Zimmerer, V. (2021). The relationship between language impairment and narrative organisation: New methods to measure deviation from the "typical structure". *59th Academy of Aphasia Annual Meeting*, EasyChair Preprint 6409. https://doi.org/10.13140/RG.2.2.12304.33280 [Funding statement: Arts and Humanities Research Council, project AH/L004070/1]

Zimmerer, V., Tsoukala, A., Cokal, D., Sevilla, G., Douglas, M., Jones, M., Ferrier, N., Turkington, D., Watson, S., Varley, R., Hinzen, W. (2024) The relationship between language disorder and thought disorder: comparing spoken narratives of people with aphasia and people with schizophrenia. *23rd Science of Aphasia Conference*, Geneva, Switzerland. https://doi.org/10.17605/OSF.IO/FX39T

7. Contacts

Please contact the MacroNarr team for any questions or issues you may have at macronarr@gmail.com.

Please note that **MacroNarr is still under development**, and while we aim to ensure a smooth and reliable user experience, **occasional issues may arise as we refine the tool**. These could include performance limitations and slower loading times, especially when running Pyodide directly in the browser, which can be affected by large datasets or varying browser support. Another issue concerns limitations with handling certain types of input (e.g. string input made up of symbols for your proposition codes, or using characters in other languages). Although differences in file encoding could be problematic, MacroNarr uses utf-8-sig with BOM for facilitating detection of UTF-8 encoding (i.e., the standard). **Please avoid ISO-8859-1 and ASCII.** We are currently working on addressing these and other potential challenges and improving functionality. If you encounter any other problems, please send us an email.

Additionally, please note that it may not always be possible to reply in a timely manner, subject to other commitments and constraints. If you have any urgent queries, you can also reach Andromachi Tsoukala at adt48@cam.ac.uk.

8. References

Knowlton, B. J., & Squire, L. R. (1994). The information acquired during artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(1), 79–91. https://doi.org/10.1037/0278-7393.20.1.79

Nicholas, L. E., & Brookshire, R. H. (1995). Presence, completeness, and accuracy of main concepts in the connected speech of non-brain-damaged adults and adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 38(1), 145-156. https://doi.org/10.1044/jshr.3801.145

Richardson, J. D., & Dalton, S. G. (2016). Main concepts for three different discourse tasks in a large non-clinical sample. *Aphasiology*, 30(1), 45-73. https://doi.org/10.1080/02687038.2015.1057891