Webinar: From Transcript to Insight
An introduction to SALT

Marleen Westerveld, PhD

Thanks for coming along! This webinar will begin shortly.

The webinar will be recorded so it is accessible to others after the live session.

Please remember to turn off your camera and microphone.
Feel free to post your questions into the chat box. I will attempt to answer them at the end of the presentation.

Follow me on Instagram for updates:
@dr.marleenwesterveld.slp

Marleen Westerveld, PhD

From Transcript to Insight – An Introduction to SALT

Overview

- Disclaimer
- Why LSA?
- What is SALT?
- SALT compared to SUGAR, CLAN, MIMO
- Choosing the sampling context
- A Case Study to demonstrate how to enter, code, and analyse a story retell sample
- Questions & Answers

Why Language Sample Analysis? all depends on your clinical question…

1. To determine performance at word- and sentence-level?
2. To determine performance at discourse level? Conversation, narration, expository, persuasion?
3. To set goals for intervention?
4. To assist in monitoring progress over time?
5. To compare results to typical peers [at word-, sentence- and/or discourse level] to confirm/refute/complement standardized test results?
6. To create performance reports for reporting back to family, teachers, other professionals?

Why does LSA feel time consuming?

- Transcription – BUT now we have ASR!
- Utterance segmentation (~ affects MLU)
- ID bound morphemes (or check them)
- Identify repetitions, unfinished utterances, reformulations (i.e. mazes)
- Identify pauses (or not!)
- Identify errors at word- and sentence-level (or not – but is a sensitive indicator of language difficulties)
- And then there is analysis / interpretation

Transcription: Automated Speech Recognition (ASR)

Try: https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/#features or www.otter.ai/ or www.rev.ai or whisper

Or dictate the child's sample into: https://dictation.io/speech

Or: Send to SALT

Or: automatically transcribe using the freely available TalkBank Batchalign (either online or locally) – more later.

Pitfalls of ASR

- May not be HIPAA compliant – client confidentiality
- Using Microsoft – does not identify mazes and fillers and often ‘corrects’ the grammar!
- Does not work well for transcripts that are difficult to understand (audio quality or intelligibility).

A safe approach atm…

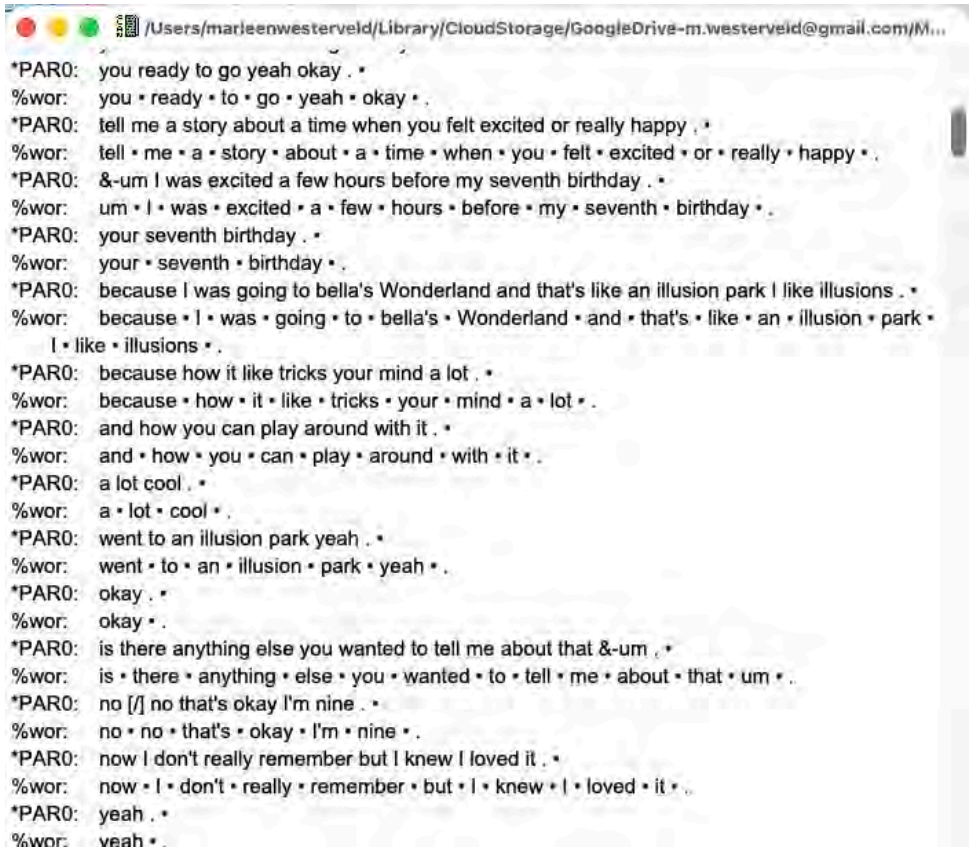
- I use Batchalign – locally with Whisper.ai - to create a CHAT file.
- Cheat-sheet for installation available or go to GitHub
- You then have 2 options

  * Use CLAN to tidy up the transcript

  * and/or convert the CHAT into a SALT ‘friendly’ file (txt) – chat2txt.marleenwesterveld.com

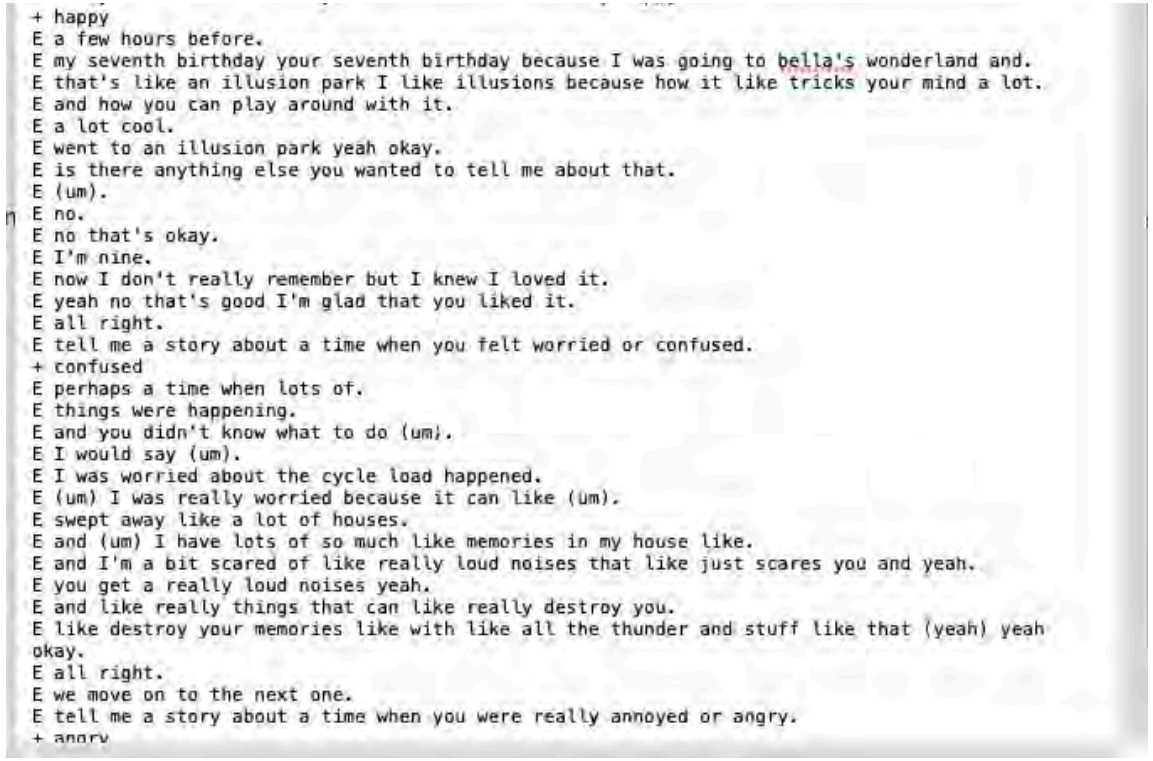
From Audio to Transcript

Step 1: Batchalign



Step 2: Chat 2 TXT;

- Tidy up speaker ID
- Add prompts (Global TALES) (+lines)
- Add timelines (stop-start)
- Tidy up the mazes.



Let’s move on to analysis tools / computer programs

Several options:

- SALT
- CLAN
- SUGAR
- MiMo

What is SALT?

Systematic Analysis of Language Transcripts (SALT) is software that:

- standardizes the process of eliciting, transcribing, and analyzing language samples.
- includes a transcription editor, standard reports, quick snapshots and reference databases for comparison with typical peers.

Available in English, Spanish and Turkish

Can handle L ➞ R languages for simple analysis – MLU, NDW, Utterances.

https://www.saltsoftware.com/products/software

SALT Built-in Reference Databases [Benchmarks]

https://www.saltsoftware.com/resources/databases

Databases of English-fluent Speakers

- Play
  - ages 2;8 - 5;8 (grades Pre-K and K)
- Conversation
  - ages 2;9 - 13;3 (grades Pre-K, K - 3, 5, 7)
- Narrative SSS (student selects story)
  - ages 5;2 - 13;3 (grades K - 3, 5, 7)
- Narrative Story Retell
  - based on 4 stories
  - ages 4;4 - 12;8 (grades Pre-K, K - 6)
- Expository
  - ages 10;7 - 18;9 (grades 5 - 7, 9 - 12)
- Persuasion
  - ages 12;10 - 18;9 (grades 9 - 12)

Databases Contributed by Colleagues

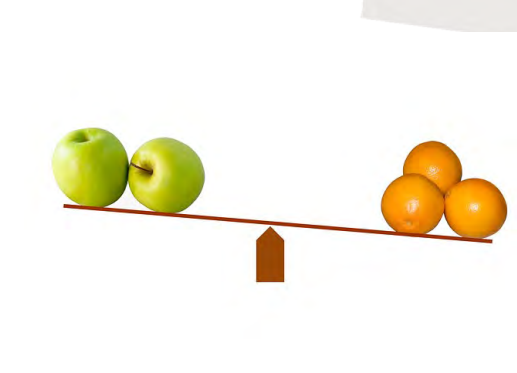
- TNL2 Narrative Samples
  - narratives from TNL-2 (Gillam & Pearson, 2017)
  - ages 4;0 – 14;11
- TNL Narrative Samples
  - narratives from TNL (Gillam & Pearson, 2004)
  - ages 5;0 – 11;11
- New Zealand/Australia databases
  - conversation, personal narrative, story retell, and expository
  - ages 4;0 - 8;9
- ENNI
  - narratives from ENNI (Schneider, Dube & Hayward, 2005)
  - ages 4;0 - 10;0

Hill et al.'s adolescent language samples: Expository, persuasion, recount, narrative (age 12 to 15)

Let’s compare SALT to….

CHAT and CLAN: http://dali. talkbank.org/clan/

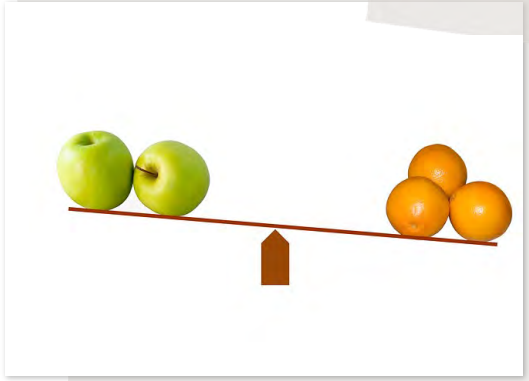
- Computes LSA measures across 49 languages
- Requires ‘coding’ for repetitions, fillers, intelligible segments, abandoned utterances, grammatical errors – BUT ASR is getting good at this!
- Can automatically check/code morphology (in many languages)
- Can link to audio and/or video ☺
- Some published measures available for MLU, VocD, IPSync, DSS < 6 years (KIDEVAL).



Let’s compare SALT to….

CHAT and CLAN:

Currently, we should not compare anything other than adult-child play with a child under the age of 6 who speaks a mainstream dialect of English to the database. Note: Clinicians are advised to use KidEval reference values in conjunction with other language performance indicators (e.g., standardized test scores). A child's performance can also be benchmarked by prior research findings for individual LSA measures (such as MLU, NDW, etc.) summarized in the SLP Guide to CLAN Appendices for further information - https://talkbank.org/0info/manuals/Clin-CLAN.pdf



You can also calculate:

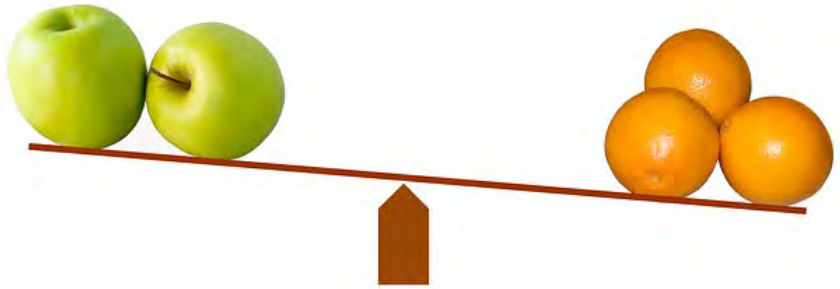
- Brown’s morphemes
- You can use EVAL to look at adult language ability.
- Fluency

Let’s compare SALT to….

SUGAR: https://www.sugarlanguage.org/ is a protocol- using MS Word.

Minimal coding required – although identification of morphemes is needed.

- Transcribers omit filler words, repetitions, and reformulations
- No procedures for error coding
- Comparison data available for 50 utterance *conversation* samples (3-7 years)
- But - conversation samples not the best context for assessing syntactic complexity

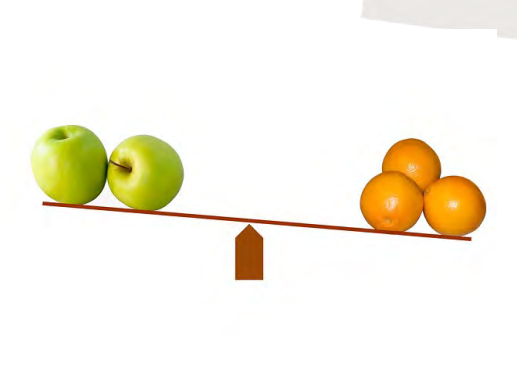


Let’s compare SALT to….

MiMo (minimal inputs, maximal outputs philosophy)

https://mimolanguageanalysis.uk/

- Grammar available for 65 mainly European languages
- Requires indication of speaker (CHI: or EXA:) similar to CHAT.
- Perhaps save CHAT files to TXT files and upload directly.
- Or use the Chat2Txt converter [more later].
- Cool colour coding of lexical and syntactic classes
- You can compare [MLU-M and MLU, HDD and T[illegible]] to norms taken from CHILDES



Example MIMO output

line	speaker	sentence_coloured	Num Words	Num Morphs	num_clause	num_fin_clause	turn_length	Morph Complex Words
1	CHI:	An na was **going** out of the door .	7	9	1	1	7	going
2	CHI:	and he **looked** into the shops .	6	8	1	1	7	shops looked
3	CHI:	and him was **crying** .	4	5	1	1	7	crying
4	CHI:	And the police pat on his back .	7	7	0	0	7	
5	CHI:	and **take** her home .	4	4	1	1	7	
6	CHI:	and him **got** lost .	4	4	1	1	7	
7	CHI:	And the policeman **said** you do n't **have** to **go** out of ever again .	13	14	3	2	7	don't

Showing 1 to 7 of 7 entries

So how long does it take? (50 utterances)

SALT (5-15mins)

C Ana get/3s sad.
C Does his[ep:he] get lost?
; :05
C His mum And dad come to fishing [eu].
C (he) His brother say[ew] wanna play?
C And look.
; :04
C Ana get/*3s lost.
C And which way will they go?
C That way.
C And never to a beach [eu].
C She went to a beach.
C X sad {crying sound}.
C And he[ew] saw a policeman.
C a policeman He said "you/'re Ana"?
C Yes.
C And yes.
C X And *the policeman took her home.
C And *he said to her "don't get lost".
=C The end.
E fantastic.
- 01:10

SUGAR (2-4 mins)

1. Use MS Word
2. Transcribe 50 utterances
3. Do not transcribe mazes
4. Divide into morphemes manually

CLAN (5 – 11mins)

Automatic morpheme coding

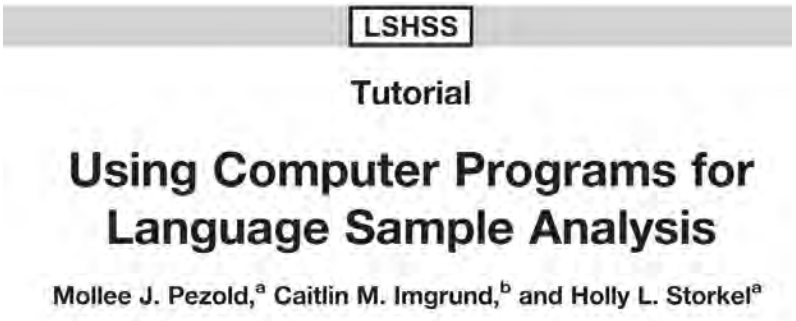
```
@Begin
@Languages: en
@Participants: CHI Child, MOT Mother
@ID: en|TalkBank|CHI|2;10.15|||Target_Child|||
@ID: en|TalkBank|MOT|||||Mother|||
@Media: samplefile, audio
@Options: CA
*MOT: what are you drawing? [b]
*CHI: a dog. [b]
%mor: det:a n:dog
*MOT: a big dog? [b]
*CHI: no, a little dog. [b]
@End
```

(Pezold et al., 2019)

It all depends on your LSA questions…

- Types of measures you are interested in: MLU, NDW, Grammatical Errors, Rate, Verbal Fluency, Turntaking, Narrative quality etc.
- Depth of your planned analysis
- Age of your client
- Intelligibility
- Discourse context
- Describing and/or comparing to benchmarks.

Comparison of database features


LSHSS

Tutorial

Using Computer Programs for Language Sample Analysis

Mollee J. Pezold,[a] Caitlin M. Imgrund,[b] and Holly L. Storkel[a]

Table 1. Comparison database features across CLAN, SALT, and SUGAR.

Feature	Category	CLAN	SALT	SUGAR
Age	Early childhood	✓	✓	
	Preschool (3–5)	✓	✓	✓
	Early school age (up to 7)		✓	✓
	Later school age (8+)		✓	
Type of sample	Play	✓	✓	
	Conversation		✓	✓
	Narrative		✓	
	Expository		✓	
	Persuasion		✓	
Languages	English	✓	✓	✓
	Bilingual English/Spanish		✓	
	Spanish		✓	
Cost	Free	✓	US$249	✓

Note. CLAN = Computerized Language Analysis; SALT = Systematic Analysis of Language Transcripts; SUGAR = Sampling Utterances and Grammatical Analysis Revised.

Analysis features

Measures	CLAN	SALT	SUGAR
MLU (m)	yes	yes	yes
CPS		yes	yes
~~TTR~~	~~yes~~	~~yes~~	~~Yes~~ ☹
Mazing		yes	x
Gramm. Acc		yes	
Examiner behaviour	Turns	yes	

Worth noting:

Miller, Andriacchi & Nockerts (2016).

Students with language disorders are not alike:

a. **General developmental delay-** low mean length of utterance, number of different words, total words and words per minute.
b. **Word finding and utterance formulation problems**- high number of repetitions and revisions at the word or phrase level.
c. **Discourse difficulties** trouble maintaining topic, failure to respond to examiner questions.
d. **Slow speaking rate-** low words per minute, high number of pauses
e. **Fast speaking rate with low semantic content-** high words per minute with circumlocution and mazing.
f. **"Disordered" language -** high percentage of errors and omissions.

Measures of fluency, narrative macrostructure, articulation and written language can also be obtained.

LSA, when done properly, is a "standardised assessment".

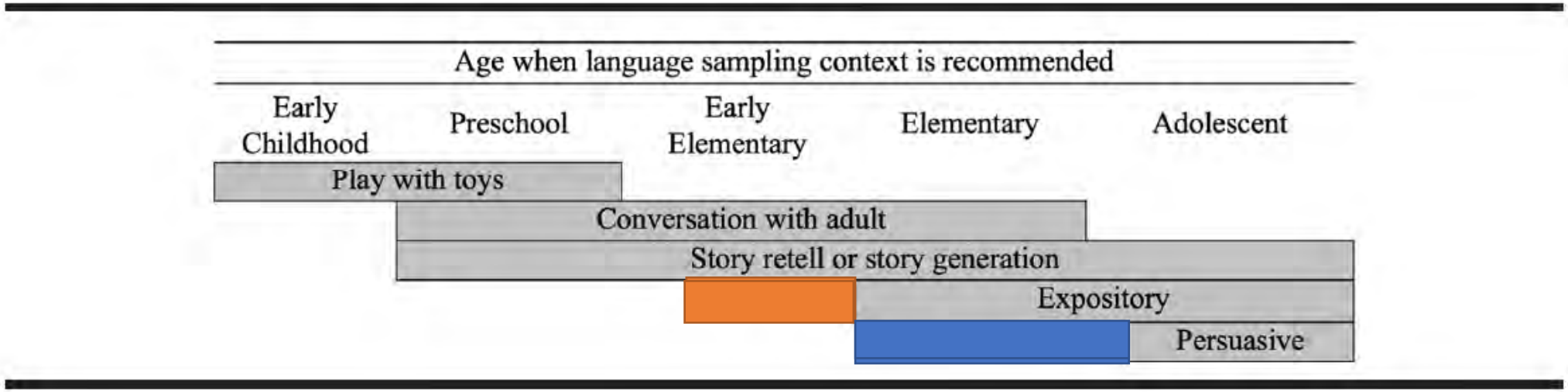
Choosing the elicitation context

Depends on:

Child's ~ MLU [≥ 3]
Child's intelligibility: 70-80%
Child's age
The reason for eliciting the sample
Child's context (home, school, community)
Child's / Family's goals
Planned analysis
Availability of 'norms' or benchmarks

A rough guide

Figure 1. Recommended language sample contexts by age.



(Pezold et al., 2020)

My suggestion

- Check why you want to do LSA.
- Do you have access to SALT?
- Do you want to compare to benchmarks?
- If so, check out the built-in reference databases as you cannot compare:

Case study - Hugo

Age 5 years, 10 months

History of speech/language difficulties

History of SLP intervention

You elicit a language sample as part of your six-monthly intervention review process.

CELF-P2 results

Hugo is in Prep; teacher has concerns about his participation in the classroom.

Word Structure:	SS 6
Expressive Vocabulary:	SS 9
Recalling Sentences:	SS 6
Sum of Expressive language scores:	SS 83
Basic Concepts:	SS 11
Concepts and Following Directions:	SS 5
Sentence Structure:	SS 8
Sum of Receptive language scores:	SS 89

Decisions

Discourse context:

* Age appropriate, relevant to curriculum:

* story retell and comprehension

* Compared to benchmarks

Hypotheses / Questions

* Difficulties with syntax/morphology in spoken discourse?

* Ability to tell a coherent story?

* Difficulties in story comprehension?

Measures:

* MLU, Grammatical accuracy / errors,

* Narrative Quality

* Narrative Comprehension

SALT Built-in Reference Databases

https://www.saltsoftware.com/resources/databases

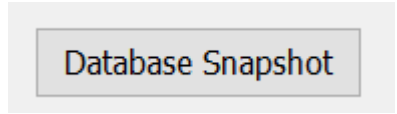
Databases of English-fluent Speakers

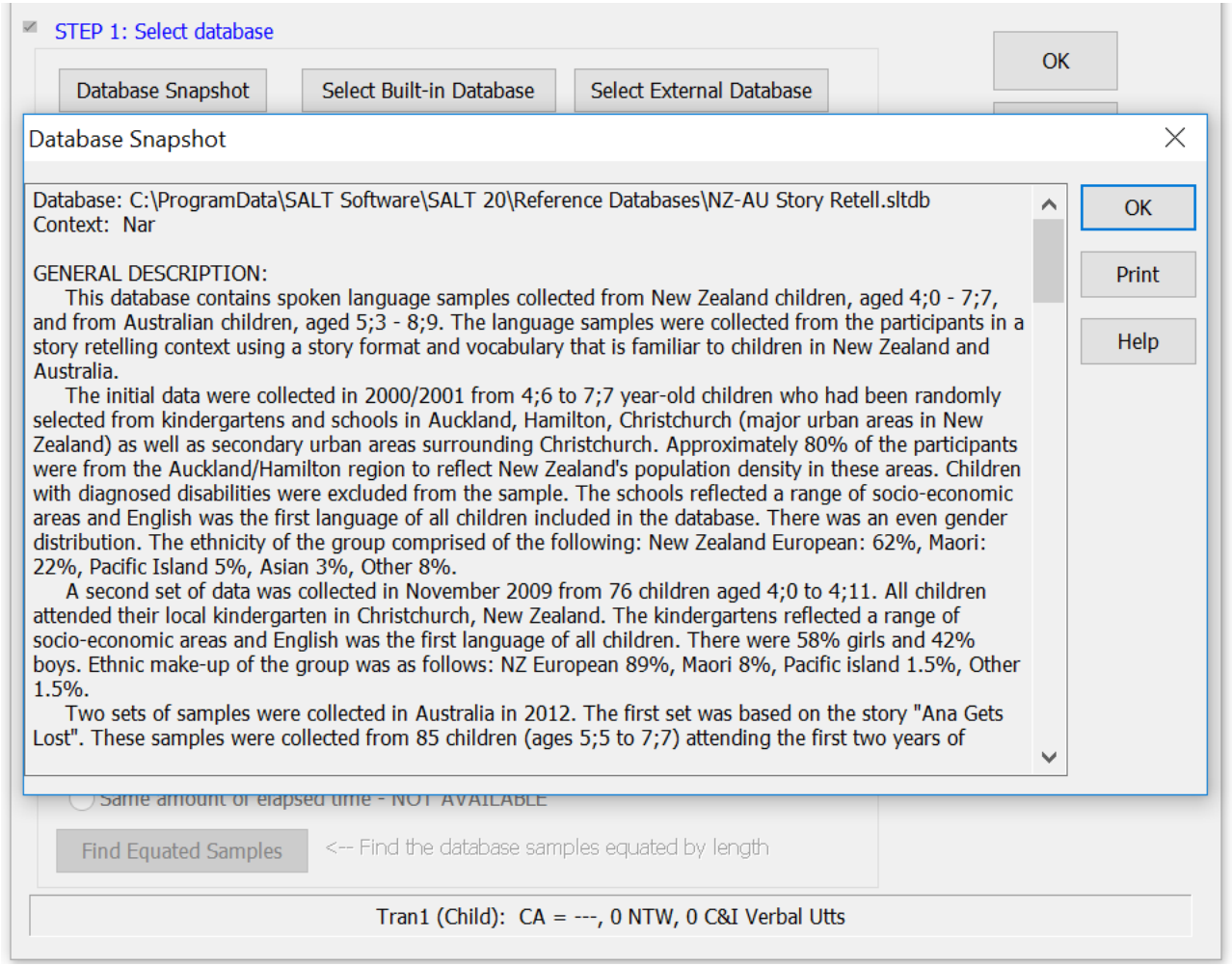
- Play
  - ages 2;8 - 5;8 (grades Pre-K and K)
- Conversation
  - ages 2;9 - 13;3 (grades Pre-K, K - 3, 5, 7)
- Narrative SSS (student selects story)
  - ages 5;2 - 13;3 (grades K - 3, 5, 7)
- Narrative Story Retell
  - based on 4 stories
  - ages 4;4 - 12;8 (grades Pre-K, K - 6)
- Expository
  - ages 10;7 - 18;9 (grades 5 - 7, 9 - 12)
- Persuasion
  - ages 12;10 - 18;9 (grades 9 - 12)

Databases Contributed by Colleagues

- TNL2 Narrative Samples
  - narratives from TNL-2 (Gillam & Pearson, 2017)
  - ages 4;0 – 14;11
- TNL Narrative Samples
  - narratives from TNL (Gillam & Pearson, 2004)
  - ages 5;0 – 11;11
- New Zealand/Australia databases
  - conversation, personal narrative, story retell, and expository
  - ages 4;0 - 8;9
- ENNI
  - narratives from ENNI (Schneider, Dube & Hayward, 2005)
  - ages 4;0 - 10;0

How to elicit the narrative?





This will outline the procedures; provide transcription notes; give information about the database location etc.

For Australia and NZ:

Ana Gets Lost: ages 4;0 – 7;7

Bus Story: Prep, Grade 1, Grade 2

Retell: Ana Gets Lost (email me for a copy of the task)

- 0:00

E what happened in the beginning?

C Anna was going out of the door.

E uhuh.

C and he looked into the shops.

C and him was crying.

E uhuh.

C And the police pat on his back.

E uhuh.

C and take her home.

E uhuh.

C and him got lost.

C And the policeman said you don't have to go out of ever again.

- 1:20

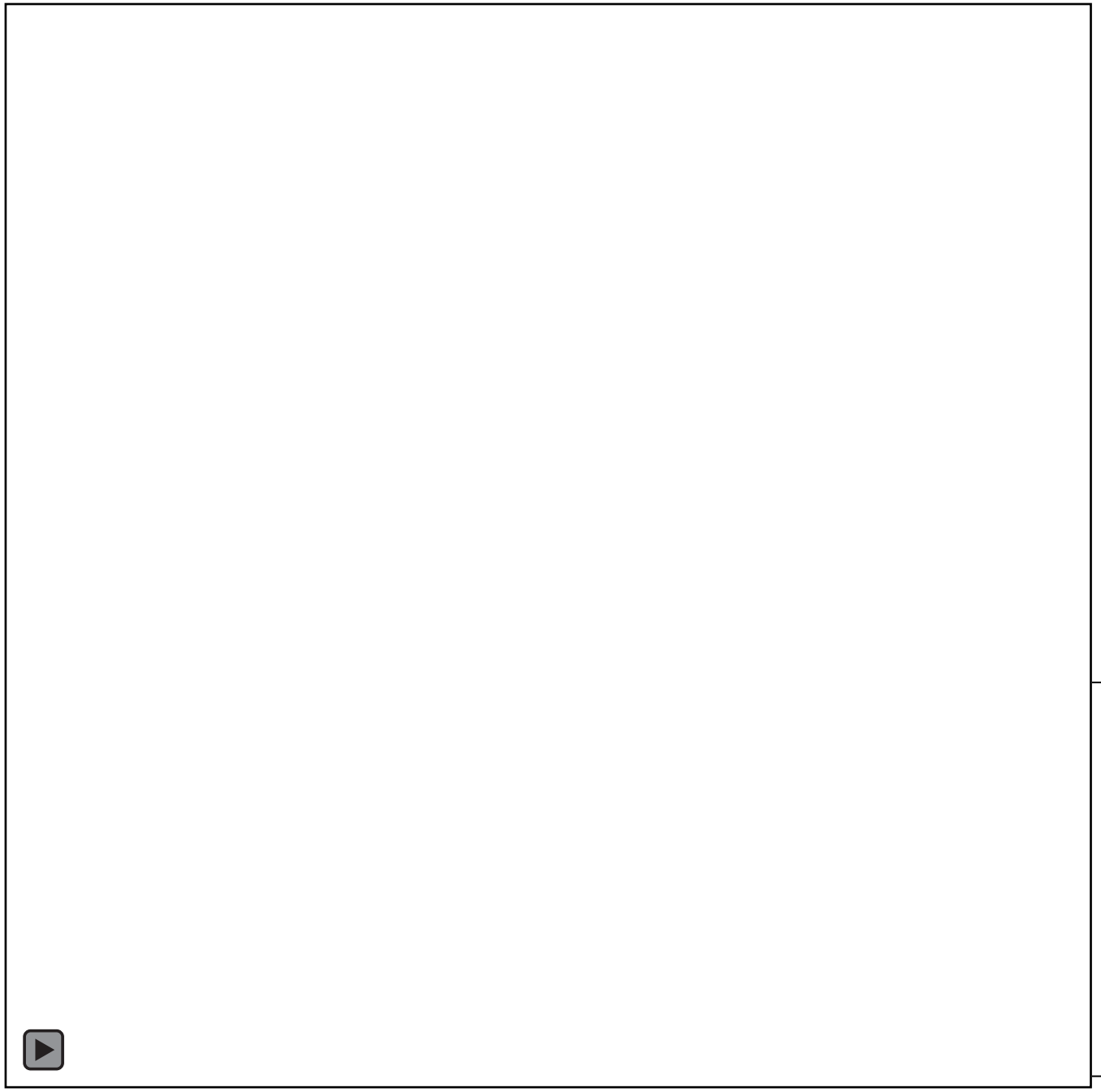
ONC: Hugo obtained a story comprehension score of 1 (out of 8)

ONQ: Hugo scored 16 on the oral narrative quality rubric.

What does this look like in SALT?

To view the demonstration video please visit YouTube: https://youtu.be/wx4WdN4PI0s

1. Start a new transcript
2. Enter some details
3. Paste / Enter the transcript (only child utterances is fine)
4. Edit – identify roots (bound morphemes)
5. Code for grammatical errors
6. Check for entry errors in SALT
7. Run - analyze



- Inserting ONC and ONQ templates
- Scoring Oral Narrative Quality

Characteristic	Content
Introduction	*Stay at home/* *couldn't go* *Parents go out* *Ana is sick* *Brother looks after her*
Theme	*Gets lost*
Main Character	*Ana*
Supporting Character/s	*Mum and dad* *big brother or (big) Brother Tom* *A Policeman*
Conflict	*Bored.* Includes rationale for character's behaviour. Ana goes out to find her parents because she is bored. Provides the relationship connecting events and actions.
Coherence	Critical Events: o *Parents have gone out.* o *Leaves the house to look for mum and dad.* o *Gets lost /or/ not know what to do and cry* o *Policeman finds her* o *Policeman takes her home*
Resolution	Resolutions: *Home (safely) Parents happy OR Parents thank the police.*
Conclusion	Endings: *Policeman told her not to get lost again AND/OR Policeman drove away OR good alternative, e.g., "mother said she wouldn't get lost again."*

Analysis

Our questions:

- MLU
- Grammatical accuracy /errors
- Story comprehension
- Story quality

To view the demonstration video please visit YouTube:
https://youtu.be/wx4WdN4PI0s

“Quick Look”

Overview of strengths and challenges across the different domains:

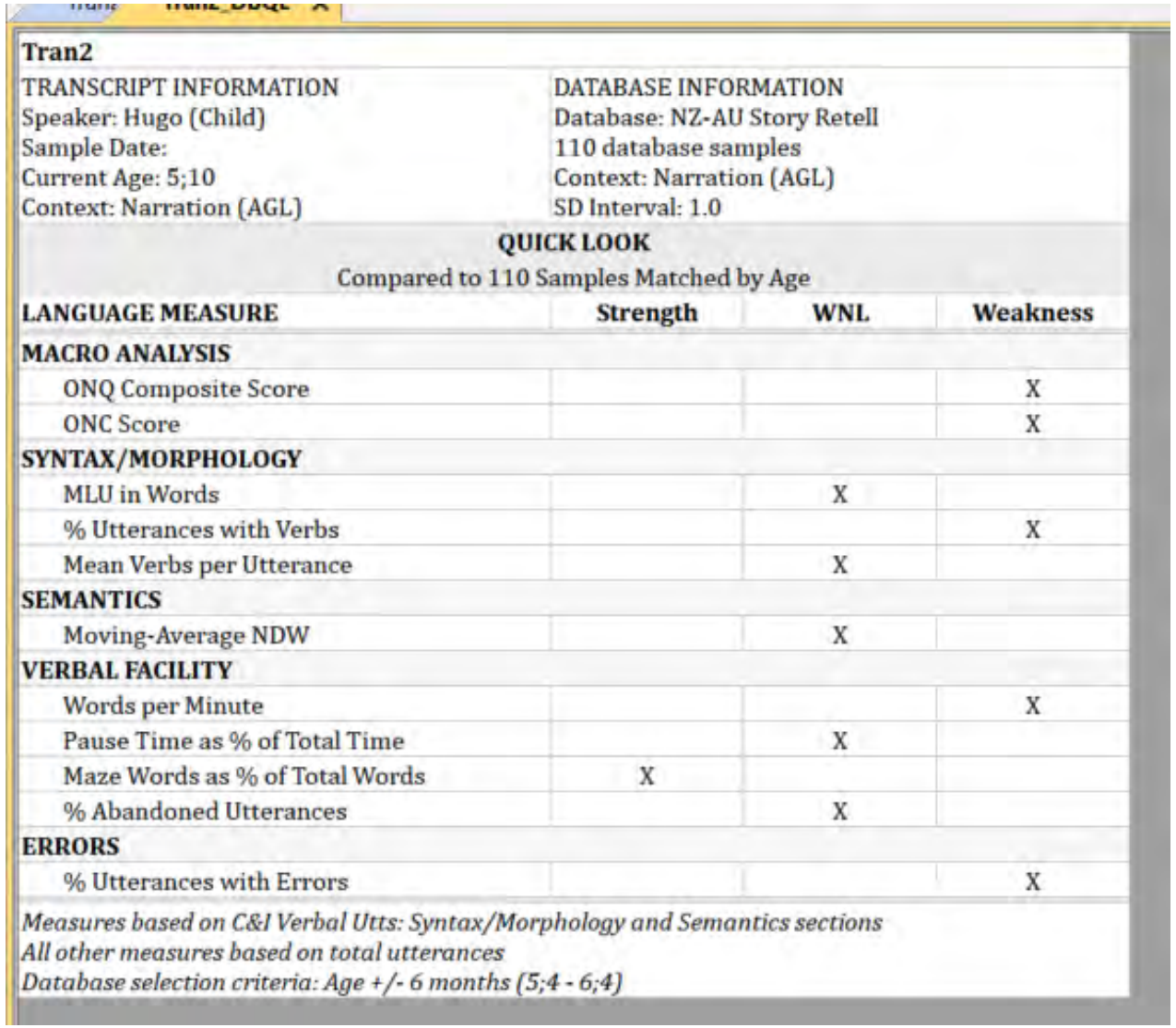
Oral Narrative quality
Oral narrative comprehension

MLU

Grammatical accuracy

Our questions:

- MLU
- Grammatical accuracy /errors
- Story comprehension
- Story quality



Tran2

TRANSCRIPT INFORMATION	DATABASE INFORMATION
Speaker: Hugo (Child)	Database: NZ-AU Story Retell
Sample Date:	110 database samples
Current Age: 5;10	Context: Narration (AGL)
Context: Narration (AGL)	SD Interval: 1.0

QUICK LOOK

Compared to 110 Samples Matched by Age

LANGUAGE MEASURE	Strength	WNL	Weakness
MACRO ANALYSIS			
ONQ Composite Score			X
ONC Score			X
SYNTAX/MORPHOLOGY			
MLU in Words		X	
% Utterances with Verbs			X
Mean Verbs per Utterance		X	
SEMANTICS			
Moving-Average NDW		X	
VERBAL FACILITY			
Words per Minute			X
Pause Time as % of Total Time		X	
Maze Words as % of Total Words	X		
% Abandoned Utterances		X	
ERRORS			
% Utterances with Errors			X

Measures based on C&I Verbal Utts: Syntax/Morphology and Semantics sections
All other measures based on total utterances
Database selection criteria: Age +/- 6 months (5;4 - 6;4)

Shaded in grey: 1 or 2 SD from the mean (* or **)

TRANSCRIPT INFORMATION
Speaker: Hugo (Child)
Sample Date:
Current Age: 5;10
Context: Narration (AGL)

DATABASE INFORMATION
Database: NZ-AU Story Retell
110/97 database samples
Context: Narration (AGL)
SD Interval: 1.0

STANDARD MEASURES REPORT
Compared to 110 Samples Matched by Age
Italicized Measures Compared to 97 Samples Equated By Same Number of Total Words

LANGUAGE MEASURE	Child		DATABASE			
	Score	+/-SD	Mean	Min	Max	SD
Current Age (5;10)	5.83	-0.34	5.95	5.33	6.33	0.34
TRANSCRIPT LENGTH						
Total Utterances	7 *	-1.21	12.61	4	25	4.63
C&I Verbal Utts	7 *	-1.12	12.08	4	23	4.54
All Words Including Mazes	45 *	-1.24	96.82	19	226	41.85
Elapsed Time (1:20)	1.33	-0.28	1.56	0.43	6.50	0.80
INTELLIGIBILITY						
% Intelligible Utterances	100%	0.31	98.87	75.00	100.00	3.72
% Intelligible Words	100%	0.32	99.82	97.06	100.00	0.56
MACRO ANALYSIS						
ONQ Composite Score	16 *	-1.22	24.16	12	40	6.68
ONC Score	1 **	-3.76	6.26	2	8	1.40
SYNTAX/MORPHOLOGY						
MLU in Words	6.43	-0.37	6.87	3.85	10.38	1.18
MLU in Morphemes	6.86	-0.34	7.29	4.23	11.13	1.26
% Utterances with Verbs	85.7% *	-1.60	95.86	75.00	100.00	6.33
Mean Verbs per Utterance	1.29	-0.66	1.47	1.00	2.78	0.28
SEMANTICS						
Number Total Words (NTW)	45	0.00	45.00	45	45	0.01
Number Different Words (NDW)	32	0.49	30.46	23	39	3.14
Moving-Average NTW	45	0.29	43.67	18	45	4.61
Moving-Average NDW	32	0.89	28.71	13	36	3.68
Moving-Average Type-Token Ratio (TTR)	0.71	0.94	0.66	0.51	0.80	0.06
VERBAL FACILITY						
Words per Minute	33.75 *	-1.38	67.45	14.15	133.33	24.38
Pause Time as % of Total Time	0.0%	-0.66	6.47	0.00	52.22	9.78
Maze Words as % of Total Words	0.0% *	-1.41	12.12	0.00	42.42	8.59
% Abandoned Utterances	0.0%	-0.51	2.73	0.00	20.00	5.35
ERRORS						
% Utterances with Errors	85.7% **	4.17	17.60	0.00	77.78	16.32
% Word Errors	8.9% **	2.57	2.43	0.00	10.17	2.51

** At least 1 SD (** 2 SD) from the database mean*
Measures based on C&I Verbal Utts: Syntax/Morphology and Semantics sections
All other measures based on total utterances
Database selection criteria: Age +/- 6 months (5;4 - 6;4)

Compared to his peers:

- Fewer utterances (Total Utterances)
- Low Oral Narrative Quality
- Low Comprehension
- Fewer utterances with verbs
- Low number of total words
- Low rate
- Low grammatical accuracy
- High number of errors

Relative strengths:

- MLU
- Number of Different Words (Semantics)
- Low mazing behaviour (no mazes!)

You can also create a performance report

Performance Report

Hugo

Age: 5;10

Language Sample Analysis with SALT Software

Elicitation Task and Database Overview

Hugo completed a narrative story retell of 'Ana gets lost' (Swan, 1992). He listened to the story once and was then asked 8 comprehension questions. He listened to the story a second time and then retold the story using his own words, without the use of the pictures. Measures of sample length, intelligibility, narrative quality, comprehension, syntax/morphology, semantics, verbal facility, and errors were calculated from his language sample and compared with samples from 110 speakers completing the same task. These speakers were within 6 months of Hugo's age. Although most measures were calculated from the entire sample, a few measures, such as total pause time and number of errors, can be affected by different sample lengths, i.e., the longer the sample, the more opportunity to produce them. For these measures, Hugo's sample was compared with a subset of 97 samples matched in length by the same number of words. All measures were interpreted using a standard deviation interval of 1.00 SD.

Transcript Length

Hugo produced 7 utterances using a total of 45 words, which were both less than his database peers completing the same task. His number of utterances and words were 1.21 SD lower and 1.24 SD lower, respectively, than the database mean. He took 1 minute and 20 seconds to complete this task, which was within normal limits.

Intelligibility

Hugo's sample was 100% intelligible.

Errors

85.7% of Hugo's utterances contained errors, which was more than 3 SD higher than the database mean. His sample contained the following extraneous words: HIM[EW] twice, HIS[EW] once, and TAKE[EW] once. His sample also contained the following 3 utterance-level errors:

C Anna was go/ing out of the door [eu].

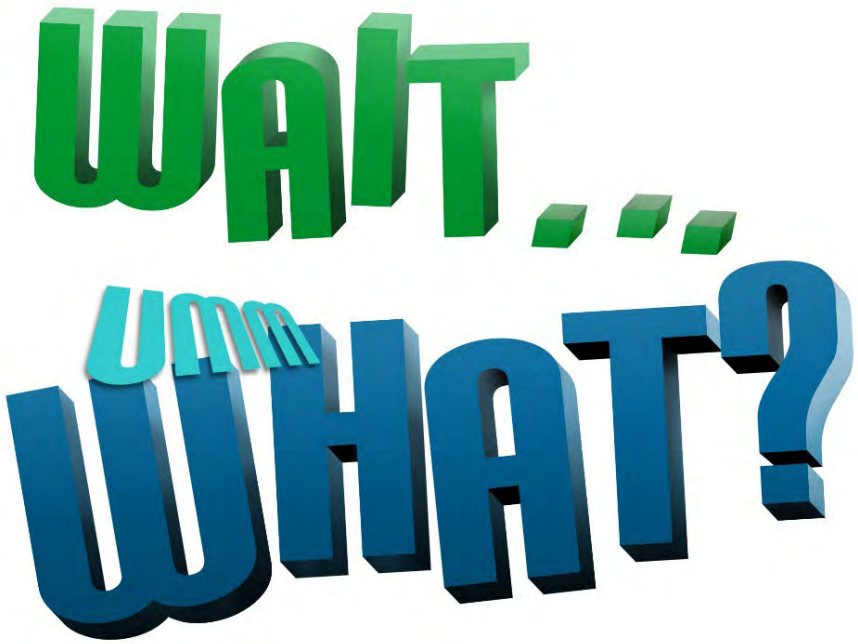
C And the police pat on his[ew] back [eu].

C And the policeman said you don't have to go out of ever again [eu].

This can be edited in Word.

But – I can transcribe and analyse 7 utterances by hand!

- That is a very short sample!!
- 50 utterances for more detailed linguistic analysis.
- This may depend on your findings from the narrative transcript analysis
- With Hugo, we also elicited personal narratives - 59 utterances, of which 17 were incomplete/unintelligible.
- And compared this to a SALT database.



Future plans

- Benchmark /reference data sets are super useful (see also Leitao et al., 2025 – Peter & the Cat [5;0 – 7;11])
- Integrating them into SALT makes it easier to compare a child's performance to their peers.
- Assistance with creating reference datasets would be super helpful.
    - E.g., Global TALES for adults!!

Future plans

- Pinches of SALT – short bite-size videos to cover the basics
  - Communication units
  - Coding for bound morphemes
  - Suggestions!!
- What (else) can I do to make language sampling and analysis more doable?

Questions/comments

You will receive an email with a link to the recording + slides
Visit www.marleenwesterveld.com – go to eLearning
Time for questions or suggestions for future webinar and/or mini-module topics!

Follow me on Instagram for updates:
@dr.marleenwesterveld.slp

Selected References

- Hill, E., Claessen, M., Whitworth, A., & Boyes, M. (2021). Profiling variability and development of spoken discourse in mainstream adolescents. *Clinical Linguistics & Phonetics*, *35*(2), 117–137. https://doi.org/10.1080/02699206.2020.1731607
- Leitão, S., Claessen, M., Visentin, D., & Calder, S. D. (2025). Narrative production in English speaking children aged 5–7 years with typical language development and developmental language disorder: Development of a reference data set. *International Journal of Speech-Language Pathology*, 1–12. https://doi.org/10.1080/17549507.2025.2536816
- Miller, J. F., Andriacchi, K., & Nockerts, A. (2016). Using language sample analysis to assess spoken language production in adolescents. *Language, speech, and hearing services in schools*, *47*(2), 99-112.
- Pavelko, S. L., & Owens, R. E. (2023). A sweet tutorial to the SUGAR method of language sampling. *Perspectives of the ASHA Special Interest Groups, 8(1), 32-49. https://doi.org/doi:10.1044/2022_PERSP-22-00134*
- Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools*, *51*(1), 103-114.