BUILDING A SPECIALISED SCHIZOPHRENIA CORPUS

Natural language approaches to the schizophrenia spectrum

OVERVIEW

Building the corpus (3 mins)

Our study results (3 mins)

Other applications of the corpus (3 mins)

Lived experience and future directions (3 mins)

BUILDING THE CORPUS

Discussing Abstract Ideas in Schizophrenia Corpus (DAIS-C)

BUILDING THE CORPUS (DAIS-C)

Discussing Abstract Ideas in Schizophrenia Corpus (DAIS-C)

Reference corpora (BNC, 2007; COCA, 2015) vs specialised corpora (Flowerdew, 2014)

Growing interest in informal conversation (CANCODE; McCarthy, 1998)

Most work has elicited FTD in controlled task contexts (Marengo et. al. 1986)

Less work has looked at FTD in fluid, informal conversation (Mikesell and Bromley, 2016) but growing interest (Batinić et al., 2021)
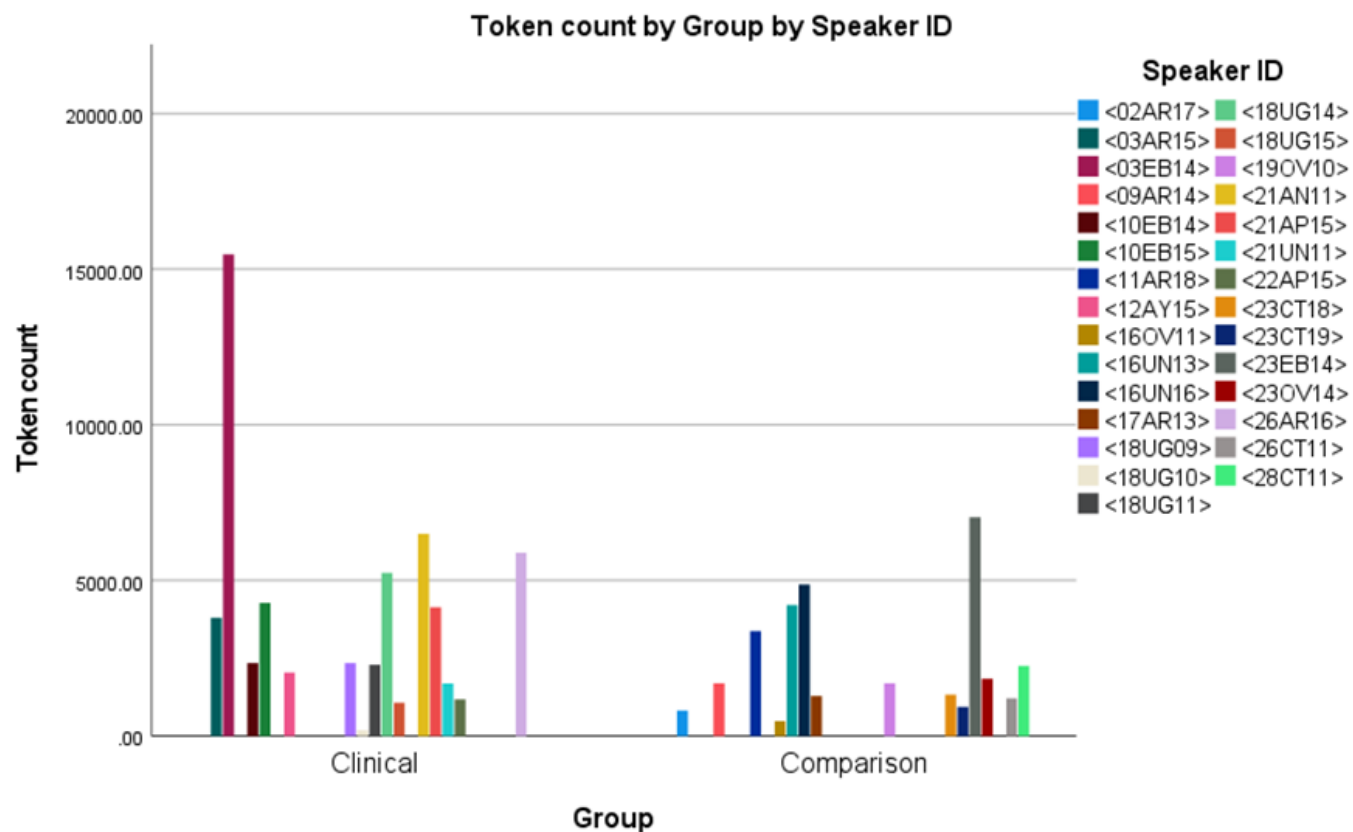
Sinclair (2005) provides comprehensive guidance on building spoken corpora

Grammatical patterns do occur with enough regularity for reliable analysis within small corpora (Carter and McCarthy, 1995)
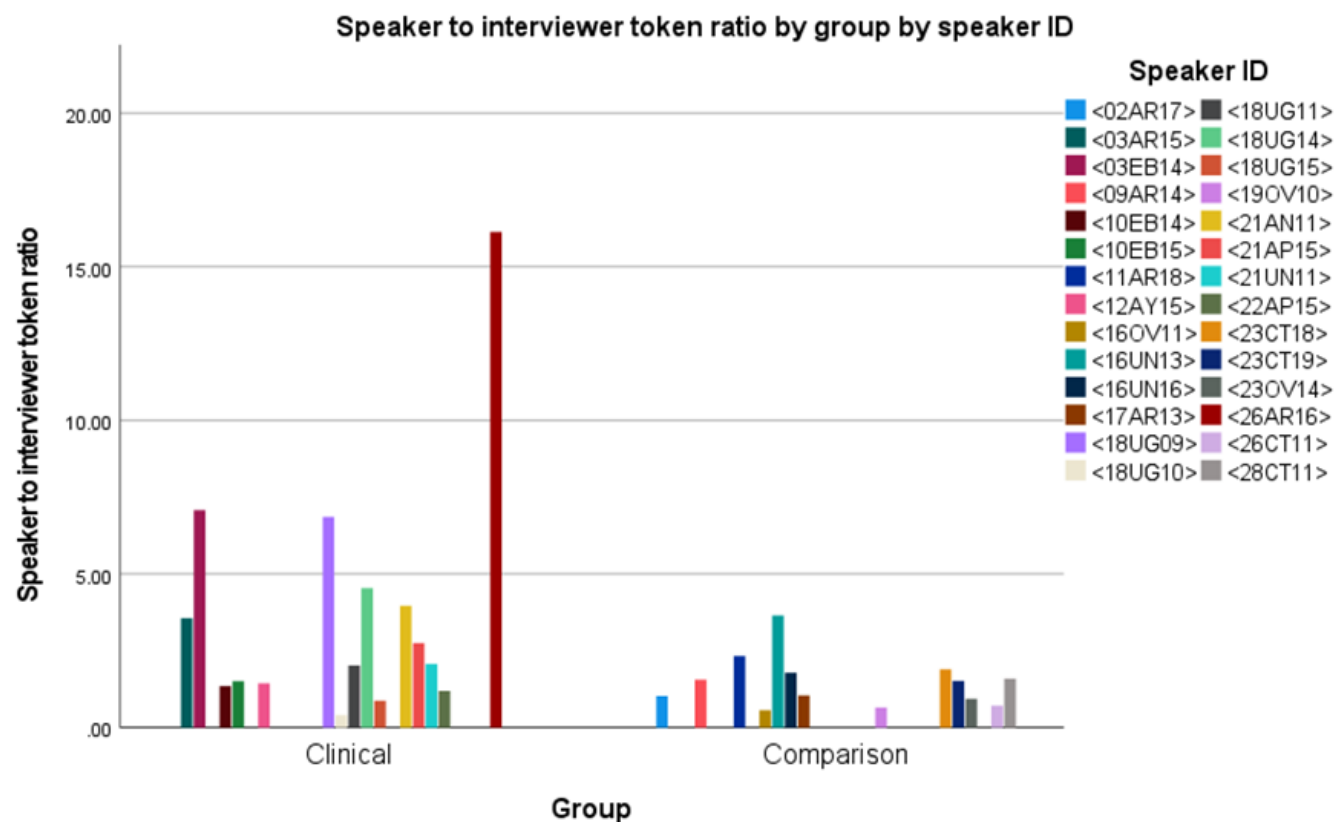
BUILDING THE CORPUS (DAIS-C)

	tokens	audio hours	audio minutes
DAIS-C	97,357	21:24:50	1284.8
DAIS-C-CL	58,444	15:47:28	947.5
DAIS-C-CO	33,025	05:37:22	337.4

BUILDING THE CORPUS (DAIS-C)

Token count by Group by Speaker ID

BUILDING THE CORPUS (DAIS-C)

Speaker to interviewer token ratio by group by speaker ID

OUR STUDY RESULTS

IRAS 225295: studies one and two

STUDY ONE: TESTING THE '4TD' FRAMEWORK

Annotated the corpus for grammatical <Gr>, word selection <WS>, thought completion <TC>, and discourse tracking <DT> features

General exclusion (very conservative annotation):

obvious self-editing and self-corrections (e.g. 'it just I'm always like I feel sad for people')

false starts (e.g. 'and you might have li I like')

**stuttering** (e.g. 'w w what')

dialectical, literacy, and EAL features (e.g. 'it sort of pops into me head'; 'I do all different places')

elision (e.g. 'like (.) trying to look at the books now there's like yeah just bits of everything really')

listing (e.g. 'animals psychology the medical stuff')

repetitions (e.g. 'yeah sensory language is a is a is a big thing')

STUDY ONE: TESTING THE '4TD' FRAMEWORK

Mean 4TD counts by historic 4TD by group
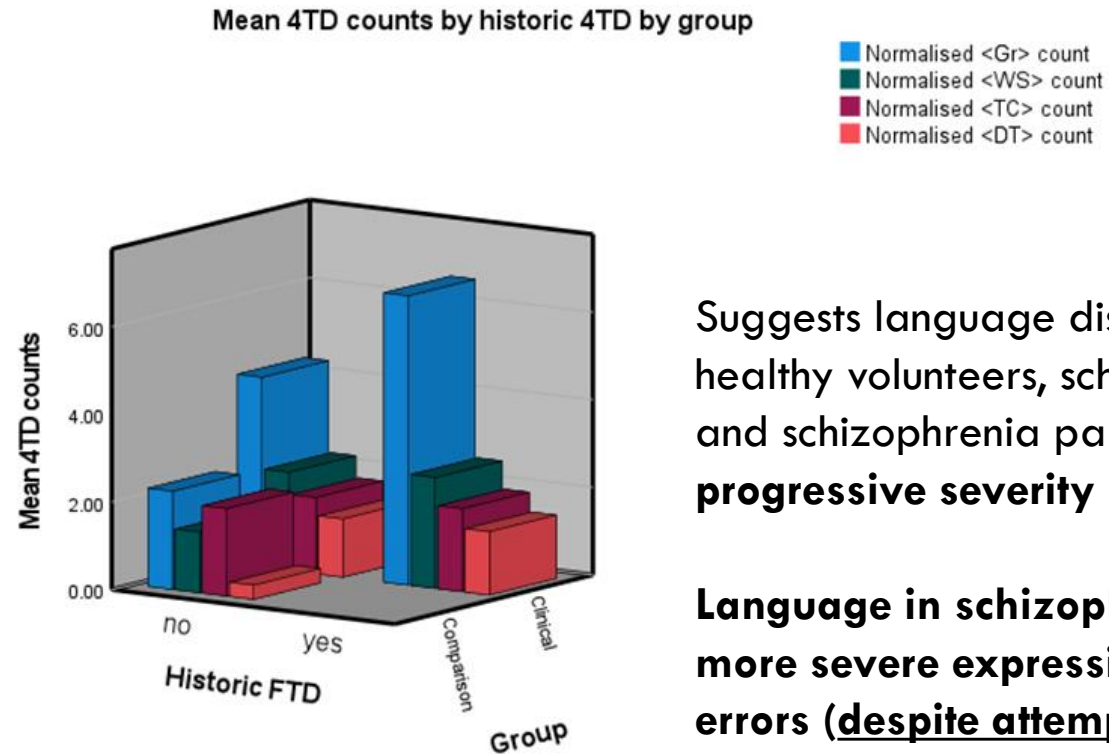
Suggests language disturbances are present in healthy volunteers, schizophrenia participants, and schizophrenia participants with FTD **in progressive severity**

Language in schizophrenia and FTD may be more severe expressions of everyday language errors (despite attempts to control for normal speech errors)

STUDY ONE: TESTING THE '4TD' FRAMEWORK

Item	O1	%1	O2	%2	LL	LogRatio
GR	570	0.94	138	0.46	64.65	1.03

$p < 0.01$

Item	O1	%1	O2	%2	LL	LogRatio
GR	570	0.94	138	0.46	64.65	1.03
DT	84	0.14	24	0.08	6.15	0.79

$p < 0.05$

STUDY ONE: TESTING THE '4TD' FRAMEWORK

Also looked at random selection of concordance lines (keyword in context/KWIC)

Organised feature types into qualitative categories

Categories were consistently similar across the subcorpora (with a slight exception in the <WS> cluster: some error types present not seen in healthy volunteers)

Type of error was less important than frequency, **suggesting language disturbances in schizophrenia are on a continuum with normal speech** (supports Thought Language Index/Liddle et al., 2018)

STUDY TWO: CONCEPTUALISING LINGUISTIC CREATIVITY

Conducted a keyness analysis (statistical test of differences across two texts)

Allowed for a comparison of statistically overused and underused terms across subcorpora

Ran keyness analysis for terms as well as semantic categories

Clinical participants referenced non-linguistic forms of creativity, such as art, computer programming, music etc. (consistent with the SLR results) and also terms unrelated to creativity, such as 'mentally ill', 'ill', 'cares', 'doctor', 'friend', 'happy', 'upset', 'jealous' 'hospital', 'prison', 'medication', 'paranoid', 'feelings', 'talk', 'difficult', 'people', and 'problem'

Semantic results echo this: 'money generally', 'greed', 'money and pay', 'unemployment', 'money: lack', 'business: selling', 'money: debts', 'clothes and personal belongings', 'quantities: little', and 'getting and possession'.

High log ratios, some terms appear up to 16 to 32 times more in the clinical subcorpus than the comparison subcorpus

Clinical speakers much more likely to depart from the interview topic and involve more topics of personal and emotional concern

SUMMARY OF FINDINGS

Study one: grammatical and discourse tracking errors significantly distinguished speaker groups (p < 0.05); language disturbances in schizophrenia appear to progress from normative error, to clinically significant, to frank FTD

Study two: clinical participants have trouble maintaining interview topic and involve topics of personal and emotional concern

SUMMARY OF FINDINGS (CONTINUED)

What could explain results?

Dysexecutive hypothesis of FTD: progressive increase in normative speech errors and difficulty maintaining interview topic could point to difficulties with executive functioning

Dyssemantic hypothesis of FTD: could explain <WS> differences and difficulty maintaining interview topic

Dysexecutive and dyssemantic are compatible, not competing, hypotheses (Dornelles and Telles-Correia, 2023)

OTHER APPLICATIONS

Discussing Abstract Ideas in Schizophrenia Corpus (DAIS-C)

CONVERSATION AND DISCOURSE ANALYSIS

Interviews presented in full transcript form

Suited to conversational and (critical) discourse analytic approaches

Can study:

relationship between interviewer and speaker

discourse markers, text moves

conversational features, paralinguistic cues (mm, er, ah)

PRAGMATIC ANALYSES

Interviews presented in full transcript form

Suited to studies of implied and other forms of referential meaning

Can study:

Gricean implicatures

metatalk, metacognition

(im)politeness, (in)civility

COMPUTATIONAL ANALYSES

Interviewee and interviewer speech isolated as plain text

Suited to algorithmic and machine-learning work

Can study:

predictive patterns across speaker subgroups

comparative analyses

texts as training data

LIVED EXPERIENCE AND FUTURE DIRECTIONS

Personal insights and areas of promise

FUTURE DIRECTIONS

Auditory verbal hallucinations, delusions, and inner speech systems

Self-monitoring failures, auditory verbal hallucinations, and FTD

From my own work:

Corpus expansion, wider (national, international, and cross-lingual) sampling

FTD and normal speech errors (tentative evidence needs further investigation)

QUESTIONS?