

# Clinical Psychology Applications of Automated Syntax Analysis

Tova Brooks  
Faculty Advisor: Mitch Marcus

## ABSTRACT

- Goal: Explore correlations between dementia related cognitive impairment and speech syntax.
- Analyze contribution of automatic syntax analysis to diagnosis of frontotemporal dementia.
- Approach: Sentence grammar structures are extracted from transcribed speech samples and analyzed for speech complexity.
- Hypothesis: patients with executive functioning deficits will use less sophisticated syntactic structures.
- Method: NLP tools to identify occurrence rates of specific sentence structures.
- Method of Analysis: Statistical tools to show significant differences between the dementia group and the controls
- Results: Showed significant differences in syntactic usage that may be explained by social-executive deficits.

## Frontotemporal Dementia - FTD

- Multiple subtypes of FTD, with different symptom clusters.
- Soc/Exec FTD: dementia affecting social comportment and executive functioning.
- Primary symptoms include inappropriate behavior, lack of insight into social interactions, decreased attention, inability to conceptually organize complex events.
- Differs from progressive non-fluent aphasia and semantic dementia, both included in the original study, which present with linguistic deficiencies.

## Data

- Cognitive task study intended to examine discourse in frontotemporal dementia patients.
- Our data is from 11 patients with Soc/Exec and 10 normal controls.
- Task: narrate an illustrated, wordless storybook as if to a child.
- These narrations were recorded and transcribed.



## Data Preparation

- Manual cleaning of dysfluencies.
- MXTerminator: identifying sentence boundaries.
- MXPOST: assigning part-of-speech tags.
- Automated parser: creates a parse tree representation of each sentence in the corpus.



## Macro view: Detectors

- Top-down method to search for relevant phenomena in the corpus.
- Detectors included the following:
  - Sentence Fragment: "The rocks and some plants."
  - Wh- questions "Where can the frog be?"
  - Yes-No questions "Is it a cat?"
  - Top level not S "Little boy and a dog, and there's a frog in a glass."
  - Fragment (FRAG tag) "Uh oh, the dog!"
  - S conjunctions "Well, he goes to bed and the frog gets out of the jar."
- Composite Detectors were used as well:
  - Leading modifiers "And the bees are chasing the dog like mad."
  - Questions
  - Sentence fragments
- A total of 14 detectors and 5 composite detectors were analyzed.

## Analysis of Detector Results

- A non-parametric Whitney-Mann U test
- The results tabulated in Figure 1 support the hypothesis

Detector	U	p <=	Lower
Leading 'But'	29	0.1	FTD
Passive Sentence	22.5	0.1	FTD
VP Conjunction	26.5	0.1	FTD
Subordinate clauses	34.5	0.1	FTD
Incomplete Sentence	21.0	0.05	Ctrl
Top Level Not S	26.5	0.1	Ctrl

These results suggest a **correlation** between lack of organizational skills and frequency of using syntax features!

## Micro view: Spine Analysis

Bottom-up method, to collect more detailed data.  
Each spine starts from the POS tag to each maximal projection.

## Analysis of Spine Results

Controls showed more proper nouns than the FTDs

More past tense.

FTDs used more demonstrative determiners ('this', 'that').

More expletive noun phrases ('there are').

More progressive, simple present tense structures.

More determiners and conjunctions.

Less passive structures.

Possible explanation:

Past tense and proper nouns creates a continuous fabric of events.

Use of -ing verbs portrays a less coherent story line.

FTD patients over-use the leading "And" is to make up for a lack of continuity in their narrative.

Altogether, this analysis indicates differing speech patterns as well!

## Classifier Training

Using machine learning to train a classifier on data of a known source and testing it on data from an unknown source.

Use cross-validation method:

Let X be the data set. Then for all x in X:

**Train** on {X - {x}}

**Test** on {x}

7 out of 11 FTD patients assigned correct label

7 out of 10 Normals assigned correct label

Classified two thirds of the data correctly! (Even with small data set!)

## Conclusions

### Success!

All three methods yielded some positive results.

This supports the hypothesis that Soc/Exec FTD may manifest in syntax as well.

However, the present data is not enough to support a reliable diagnostic tools

A larger data set is expected to support the hypothesis even more.

*Special thanks:*

*Dr. Mitch Marcus, Dr. Max Mintz and Ryan Gabbard for all their help.*

