Objective and Rationale

Treatment evaluation and research requires discourse measures that are: VALID, REUSABLE, and RELEVANT but also PRACTICAL in terms of labor and training.

Primary limiting factors for linguistic discourse analysis are:

TIME and CLINICAL KNOWLEDGE.

This project describes an automated system for measuring grammatical complexity in discourse using GRASP.

The system uses a grammatical relations (GR) parser that has recently been trained on adult language samples and tested for computation of a syntactic complexity index. By identifying the GRs that mark embedding, a Grammatical Relations-Complexity (GR-C) measure can be calculated. This GR-C measure adds important and relevant linguistic data that can be computed on multiple samples with accuracy, replicability, speed, and flexibility.

Background

Linguistic analyses in aphasia often focus on grammatical aspect because syntax is a key component of aphasia diagnosis and treatment. Many measures can be computed automatically—e.g., total utterances, total words, total unique words, TTR, MLU, words per minute, frequencies of parts-of-speech, morphological affixes, proposition density, repetitions, negatives. However, grammatical complexity has been less amenable to automatic computation.

Embedding, or recursion, is a structural indicator considered to be a good indicator of syntactic complexity. Systems used more commonly for child language, such as LARS, use embeddings to compute grammatical complexity.

GRASP (Grammatical Relations Analyzer for Spontaneous Protocols) is a parser developed to accurately and automatically measure syntactic complexity by producing a tier for grammatical relations (GRs) in CHAT files.

Recent training of the GRASP parser on adult language samples allows us to compute grammatical complexity from the GR codes.

All of the GRs (n=45) are explained in detail in the CLAN Manual (MacWhinney, 1997).

Running the MOR command in CLAN automatically produces a %mor tier with morphosyntactic analysis and a tier with pairwise GRs in words in a matter of seconds. For example: %PAR: the big dog chased five cats.

XCOMP: the non-final complex complement of a verb I thought that the cat was gray.

XCPRED: the non-final complex complement of a verb that the cat eats a cow.

XPRED: a full clause that serves as the predicate nominal of verbs. My goal is to win the competition.

PCOB: a full clause that serves as the object of a preposition. I’m not clear on why she did that.

COBI: a full clause that serves as the direct object. I remember what you said.


CDEF: a non-final clause that attaches to a verb, adjective, or adverb. We spent the day visiting museums and galleries.

NCT: the head of a complex noun phrase with a prepositional phrase attached or an adjunct of noun.

NCT: the head of a complex noun phrase with a prepositional phrase attached or an adjunct of noun.

The police saw the guy with a revolver.

CMMO: a finite clause that is a nominal modifier or complement. He was happy he found the girl.

XMOD: a non-final clause that is a nominal modifier or complement. It’s time to take a nap.

Examples of other non-embedding GRs are: SUBJ (subject), OBJ (object), PRED (predicate nominal or predicate adjective), POSS (possessive), AUX (auxiliary), NEG (negation), INF (infinitive), COM (communicator).

Testing this automated GRASP system with human coding of embeddings in adult discourse samples yielded an overall accuracy of 95% (of the 74 embeddings spotted by the automated system, two were false alarms and one was missed).

Research Questions

Will the Grammatical Relations-Complex (GR-C) measure reveal differences between Cerebral narratives produced by people with nonfluent aphasia (Broca’s), fluent aphasia (Agrammatic), and people without aphasia?

Does the GR-C Index correlate with other measures used as indicators of grammatical complexity — MLU and # of verbs/utterance?

Research Methods

Participants

Using the AphasiaBank database, inclusion criteria were:

• native English speakers
• aphasia as a result of stroke
• ≥ 20 words on the CED-R test

The test session only, if participant was seen multiple times

Demographic characteristics — means and standard deviations

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>MLU</th>
<th>Broca (n=57)</th>
<th>Controls (n=140)</th>
</tr>
</thead>
<tbody>
<tr>
<td>62.5 (12.2)</td>
<td>2.85 (0.23)</td>
<td>7.26 (1.21)</td>
<td></td>
</tr>
<tr>
<td>Education (yrs)</td>
<td>3.0 (1.8)</td>
<td>2.9 (1.74)</td>
<td></td>
</tr>
<tr>
<td>Sex (males)</td>
<td>57.9%</td>
<td>56.5%</td>
<td></td>
</tr>
<tr>
<td>WAB-R AQ</td>
<td>103.0 (9.3)</td>
<td>107.0 (9.7)</td>
<td></td>
</tr>
</tbody>
</table>

No significant group differences on age, sex, education.

Psychometric properties

Transcriptions were completed by trained and experienced transcribers. Utterances were segmented using the QPA hierarchy, syntax, intonation, pause, semantics. Two transcribers reviewed each transcription and reached forced agreement on all decisions. Complete transcripts and videos are available at http://aphasia.talkbank.org.

1. Run MOR on all transcriptions
2. Run freq nGra *PAR tag* +v2.0.10 spreadsheet of GRs % GR Complexity = (embedding GRs + total number of GRs) / 100

Of the 41 possible grammatical relations (excluding 4 cosmetic punctuation marks), the 30 that mark syntactic embeddings are:

• COMP: the finite clausal complement of a verb I thought that the cat was gray.

• XCOMP: the non-final clausal complement of a verb. It was a cow from the competition.

• CPRED: a full clause that serves as the predicate nominal of verbs. My goal is to win the competition.

• PCOB: a full clause that serves as the object of a preposition. I’m not clear on why she did that.

• COBI: a full clause that serves as the direct object. I remember what you said.

• CSTUB: the non-final complex subject of another clause. That Eric cried much.

• CDEF: a non-final clause that attaches to a verb, adjective, or adverb. We spent the day visiting museums and galleries.

• NCT: the head of a complex noun phrase with a prepositional phrase attached or an adjunct of noun.

• NCT: the head of a complex noun phrase with a prepositional phrase attached or an adjunct of noun.

The police saw the guy with a revolver.

• CMMO: a finite clause that is a nominal modifier or complement. He was happy he found the girl.

• XMOD: a non-final clause that is a nominal modifier or complement. It’s time to take a nap.

Examples of other non-embedding GRs are: SUBJ (subject), OBJ (object), PRED (predicate nominal or predicate adjective), POSS (possessive), AUX (auxiliary), NEG (negation), INF (infinitive), COM (communicator).

Results

<table>
<thead>
<tr>
<th>GR-Complexity</th>
<th>Anomic (n=87)</th>
<th>Broca (n=57)</th>
<th>Controls (n=140)</th>
</tr>
</thead>
<tbody>
<tr>
<td># verbs/utterance</td>
<td>3.9 (1.5)</td>
<td>9.9 (2.3)</td>
<td></td>
</tr>
<tr>
<td>Correlations between GR-Complex and # of verbs per utterance</td>
<td>0.69 (0.34)</td>
<td>1.0 (0.0)</td>
<td></td>
</tr>
</tbody>
</table>

Total utterances*

<table>
<thead>
<tr>
<th>MLU</th>
<th>Broca (n=57)</th>
<th>Controls (n=140)</th>
</tr>
</thead>
<tbody>
<tr>
<td>226.3 (174.6)</td>
<td>473.3 (282.3)</td>
<td></td>
</tr>
</tbody>
</table>

* significant differences among all groups (p<0.01), according to Pearson product moment correlation.

Conclusions and Future Directions

• The GR-complex measure is a promising and practical tool for automatic analysis of syntax in clinical and research discourse analyses.

• Once a discourse sample is transcribed in CHAT format, this complexity index can be automatically calculated with no further coding or annotation.

• These results are consistent with the literature, showing more embeddings in language samples of non-aphasic participants than PWA and more general findings of reduced syntactic complexity in Broca’s aphasia.

• Use the GR-complex measure:

1. to evaluate the effect of syntax treatment programs on discourse; and
2. with other populations such as Primary Progressive Aphasia, where recent neuromaging results showed that reduced embedding was associated with atrophy in the left frontal lobe (posterior inferior frontal gyrus, superior frontal sulcus and adjacent prefrontal regions and the supplementary motor area) and
3. to learn more about the types of embeddings used in different discourse genres and in participants with different types and severities of aphasia.

Ideas for improving the GR measure include the following:

• Address the fact that CSUB and COBI were not coded properly and CPRED and CSTUB were coded too infrequently on the CoRa tier;

• Consider the new Google Universal Dependency Relations initiative and consider using the relativizer as the head of the subordinate clause and simplifying the CoRa tier coding (use SUBJ and OBJ instead of CSUB and COBI) and do cross-tier searches (for example, for relativizers on the CoRa tier linked to the verb as SUBJ on the CoRa tier);

• Do additional training of the CoRa tier with more discourse genres and more samples.

References


Acknowledgments

This work was funded by NIH/NIH grant R01-DC008524 (2012-2017).